UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

Gestión Avanzada de Turnos para la Interacción Natural

Autor:
David del Valle Agudo

Directores:
Dr. Francisco Javier Calle Gómez
Dra. María Dolores Cuadra Fernández

DEPARTAMENTO DE INFORMÁTICA

Leganés, 2012

TESIS DOCTORAL

Gestión Avanzada de Turnos para la Interacción Natural

Autor: David del Valle Agudo

Directores: Dr. Francisco Javier Calle Gómez
Dra. María Dolores Cuadra Fernández

Firma del Tribunal Calificador:

Firma

Presidente:

Vocal:

Vocal:

Vocal:

Secretario:

Calificación:

Leganés, de de

Agradecimientos

Esta tesis es el resumen de un periodo de mi vida que comenzó hace ya ocho años, cuando Julio Villena me descubrió el mundo de los agentes conversacionales y de aquellas tecnologías fascinantes que hacían posible que las máquinas "hablaran". Gracias, Julio, por haberme despertado aquel interés y por haberme puesto en contacto con Paloma Martínez y con todo el Grupo LaBDA, la gente que me dio la oportunidad de cursar este doctorado y de convertir durante varios años lo que para mí no dejaba de ser una afición en mi propio trabajo. De LaBDA, y por extensión de toda SINTONIA, me llevo muchas cosas aprendidas y también un buen puñado de amigos. Merecen una mención especial Javier Calle y Dolores Cuadra, mis directores de tesis, y también otros tantos compañeros que directa o indirectamente, incluso sin saberlo, han sido parte fundamental de este trabajo y me han dado, desde un discreto segundo plano, las energías necesarias para seguir adelante. De estos últimos, y no sois pocos, guardo un especial recuerdo de los amigos del comedor. Habéis hecho inolvidables todos estos años.

También quiero agradecer a Jan Alexandersson y a todo su equipo el haberme acogido en Saarbrücken, durante la primavera y el verano de 2009, dándome la oportunidad de conocer y participar en un grupo tan destacado en el campo de la Interacción Hombre-Máquina como lo es el Departamento de Interfaces de Usuario Inteligentes del DFKI. Gracias por todo lo que me enseñasteis, y también por los buenos ratos que pasamos.

Por otro lado, sin el apoyo de las redes MAVIR (S-0505/TIC-0267) y MA2VICMR (S2009/TIC-1542); y de los proyectos IntegraTV4ALL (FIT-350301-2004-2), SOPAT (CIT-410000-2007-12), THUBAN y SemAnts (AVANZA I+D TSI-020110-2009-419) esta tesis no podría haber visto la luz. Tampoco sin la confianza depositada en esta investigación por la revista Interacting with Computer, así como por otra quincena de revistas y congresos de ámbito nacional e internacional. Gracias, también, a los doctores Anabel Gutierrez, del Regent's College de Londres, y Eric San Juan, de la Universidad de Avignon, por los informes emitidos sobre este trabajo.

Gracias, en general, a la Universidad Carlos III de Madrid y a su Departamento de Informática por haberme apoyado en los momentos buenos. Gracias, en particular, a quienes también lo hicieron en los momentos malos. En especial a Jessica, que confió en mí incluso durante todo aquel tiempo en que este trabajo, estando ya completamente terminado, tuvo que permanecer abandonado en un cajón a la espera del JCR que nunca llegaba. Aquella espera cambió mi vida. Aquella espera me hizo perder la esperanza en que este trabajo llegase algún día a ver la luz. Pero, a pesar de todo, ella siempre estuvo ahí. Ella siempre creyó en él. Ella, mi familia y también mis amigos.

ABSTRACT

As technology spreads to the different aspects of our life and it becomes more and more complex, we demand that this technology become more transparent and accessible to the people. It is for this reason that the called Natural Interaction Systems, framed in the *human-computer interaction*, are becoming of more importance in the last years. These systems try to make technology accessible through the same codes, modalities and procedures that people use when interacting with one another, without the need of applying neither previous knowledge nor specific technological abilities. In short, they seek to imitate human behaviour through a *natural interaction*.

Natural Interaction is structured in three organizational levels: *global*, which describes the links that exist between the different goals of the interaction; *local*, that represents the internal development of each goal; and *temporal*, also known as turn-taking. From among all of them, temporal organization of the interaction has been barely dealt up to this moment in the field of Natural Interaction Systems. These systems usually simplify turn-taking to a pass-the-baton process, considering it a collaborative process where two parties, user and system, alternately act in a non-overlapped way. The turn to intervene is passed from one participant to another in an organized manner, and only the one who has gained the floor can contribute to the interaction, and determine unilaterally how he will do so. In short, interaction develops from the beginning to the end as a cyclical process (that we call *interaction cycle*).

For certain interaction domains, for example some transactional domains, this could be a valid approach. However, as technology evolution and potential users demand enhanced interactive abilities (particularly pro-active capabilities and representation of sociolinguistic circumstances), such an approach seems mechanical and unnatural. The current state of the art in Natural Interaction Systems reflects several limitations related with the turn-taking developed

in the interaction. Some of these limitations are: solely bipartite interactions; utterances interpreted and generated as indivisible elements (blocking the possibility of develop an incremental processing); no handling neither overlapped turns, nor interruptions and disfluencies (phenomena of utmost importance and frequency in the development of a natural interaction); and interventions produced just as consequence of a previous turn or an internal event (without an evaluation of the opportunity and necessity of generate turns considering the rules that governs turn-taking in *human interaction*).

This work starts from the most accepted theories that describe how turn-taking is developed in the human interaction and applies them in the analysis, definition, implementation and evaluation of a set of new knowledge models that enable these systems to participate in interactions that are developed under a more human turn-taking strategy, where the order of participants' interventions are not defined beforehand, the lengths and contents of the contributions are not defined unilaterally by the participant who produces them, and where the set of possible types of contribution are not limited to primary contributions, produced only under the possession of the floor.

In short, it is attempted to make the system an active participant in the sharing out of turns in the interaction. With this aim, turn-taking is considered as a *joint action* where each participant realize his own turn-taking decision at each moment, taking into account a set of conjectures related with: the state of the goals; the commitment reached by the participants on them; the sociolinguistic circumstances that surround the interaction; the possession of floor; and finally the candidates to take it. It is tried a turn-taking decision which could even affect the ongoing system's contributions, that, far of being developed following a rigid preliminary formalization, could be reformulated (and even interrupted) during its generation.

In order to solve these problems, Natural Interaction System architectures are revised, especially with regard to the Presentation and Interaction Managers, the components that are responsible of these new turn-taking skills. These skills are: *incremental interpretation* of users' contributions; *incremental generation* of system's contributions; *turn-taking decision*; estimation of *turn-taking state* (turns state, possession of floor and candidates to take it); and *goals management* in the interaction.

The research is completed with the implementation and evaluation of a prototype of Natural Interaction System that includes the proposed knowledge models. It is presented an evaluating methodology that makes it possible to compare different turn-taking settings on the same Natural Interaction System. The compared settings are: interaction cycle (turns are developed sequentially in a predefined order of participation); advanced turn-taking (system

develops the strategies proposed in this work); and Wizard of Oz (a human determines what the system has to say and how it has to behave at all times). We put especial emphasis on the evaluation of the described skills independently of the components that are in charge of the speech acquisition, voice synthesis and natural language processing. With this aim, we define two new roles in the interaction, together with user and system. These roles are played by human participants and consist of acting as the system's input and output interfaces. This evaluation considers both technical and objective parameters gathered during the execution of the system, and the user subjective perception of its naturalness.

The results reveal improvements in both, objective metrics of interaction performance and users' subjective assessment, with respect to the interaction cycle (classic approach). It reaches very close marks to human interactive abilities in aspects such as organization of the interaction, user's preference, convenience and suitability to advanced turn-taking domains, and shows high similarity to human interaction procedures and strategies.

RESUMEN

A medida que la tecnología gana en complejidad y se extiende a más aspectos de nuestras vidas, se requiere de ella que se haga, a la vez, más transparente y accesible para las personas. Es por ello que cada vez toman mayor relevancia los denominados Sistemas de Interacción Natural, enmarcados en la *interacción hombre máquina*, cuyo objetivo es el de hacer accesible la tecnología a los usuarios a través de los mismos modos, códigos y procedimientos que los humanos utilizan de forma natural para comunicarse entre sí, sin requerir conocimientos previos o habilidades tecnológicas específicas, es decir, según una *interacción natural*.

Este propósito afecta a todos los niveles en los que se estructura el diálogo: *organización local*, *organización global* y *organización temporal*. De todos ellos, es este último nivel el que menos ha sido tratado hasta el momento en el ámbito de los Sistemas de Interacción Natural. Por lo general, en este tipo de sistemas la toma de turno tiende a ser simplificada a un proceso de paso de testigo, en el que los participantes contribuyen a la interacción en un orden predefinido e invariable y en el que es el participante en posesión de la palabra quien decide de qué contenidos dotará a su contribución y durante cuánto tiempo la desarrollará. Esta rigidez en la toma de turno es la causa de alguna de las principales limitaciones que pueden apreciarse en los actuales Sistemas de Interacción Natural: interacciones exclusivamente bipartitas; enunciaciones interpretadas y generadas como elementos indivisibles en el tiempo (sin posibilidades de evolución incremental); omisión del tratamiento de turnos solapados, interrupciones y disfluencias (de gran frecuencia de aparición e importancia en el correcto desarrollo de la interacción natural); o intervenciones producidas como mera consecuencia de los turnos previos o de eventos internos (sin tomar en consideración la oportunidad y necesidad de producir turnos según las reglas de la toma de turno que rigen la *interacción humana*).

Este trabajo parte de las teorías más aceptadas sobre la forma en la que se desarrolla la toma de turno en la interacción humana para analizar, definir, implementar y evaluar nuevos modelos de conocimiento que permitan a estos sistemas participar en interacciones desarrolladas bajo una toma de turno más humana, donde el orden de intervención de los participantes no esté definido de antemano, la duración de las contribuciones no quede determinada unilateralmente por el participante que la produce, y donde las posibles formas de contribución no queden restringidas a intervenciones primarias producidas exclusivamente bajo la posesión de la palabra.

En definitiva, se pretende hacer al sistema parte activa en el reparto de los turnos de la interacción, partiendo de una concepción de la toma de turno como una *acción combinada* en la que cada participante desarrolla su decisión de toma de turno en cada instante y en función de: sus conjeturas sobre el estado de las metas; el compromiso alcanzado entre los participantes sobre ellas; las circunstancias sociolingüísticas en las que se desarrolla la interacción; y el estado de la toma de turno (qué turnos que se están desarrollando, quién ostenta la posesión de la palabra y quiénes son los candidatos a tomarla). Una decisión de toma de turno que pueda incluso afectar a las contribuciones en curso del sistema, que lejos de desarrollase según una rígida formalización preeliminar, podrán ser reformuladas (e incluso interrumpidas) durante su generación.

Resolver estos problemas pasa por revisar las arquitecturas de los Sistemas de Interacción Natural, especialmente en lo que respecta a los componentes Gestor de Presentación y Gestor de Interacción, sobre quienes recaen las nuevas habilidades de toma de turno. Estas habilidades serán: *interpretación incremental* de las contribuciones de los interlocutores; *generación incremental* de las enunciaciones del sistema; *decisión de toma de turno*; estimación del *estado de los turnos*, estimación del estado de *posesión de la palabra*; estimación de los *participantes designados o candidatos a tomarla*; y *gestión del estado de las metas* de la interacción.

El trabajo se completa con la implementación y evaluación de un prototipo que incluye los modelos de conocimiento propuestos. Se presenta una metodología de evaluación que permite comparar diferentes configuraciones de toma de turno sobre un mismo sistema de interacción. Las configuraciones analizadas son: toma de turno por *ciclo de interacción*; toma de turno según la estrategia descrita en este trabajo; y toma de turno humana. Se hace especial hincapié en evaluar las habilidades descritas con independencia de los componentes de adquisición, síntesis y procesamiento de lenguaje natural, para lo cual se disponen participantes humanos desempeñando las funciones de interfaz de entrada y salida entre el sistema y el

usuario. La evaluación considera, tanto parámetros técnicos objetivos de la eficiencia, eficacia y calidad del funcionamiento del sistema, como la valoración subjetiva de la naturalidad de la interacción percibida por el usuario.

Los resultados revelan una elevada naturalidad de esta estrategia de toma de turno desarrollada por la propuesta de Sistema de Interacción Natural en aspectos como el orden en el desarrollo de la interacción, la preferencia de los usuarios y lo adecuada y cómoda que resulta la estrategia en situaciones de toma de turno avanzada. Del mismo modo, se consigue una importante mejora con respecto a la toma de turno por ciclo de interacción.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

ÍNDICE DE EJEMPLOS

Capítulo 1 INTRODUCCIÓN

En el campo de la *interacción hombre-máquina* (*HCI*) cada vez toma mayor relevancia la investigación en sistemas cuya meta es poner la tecnología a disposición de los usuarios a través de los mismos modos, códigos y procedimientos que los humanos utilizan entre ellos para comunicarse de forma natural, sin requerir de las personas conocimientos previos o habilidades tecnológicas específicas. El objetivo último es desarrollar sistemas que cualquier usuario pueda manejar sin contar con más herramientas que su capacidad para interactuar con otras personas, en definitiva, sistemas capaces de imitar el comportamiento interactivo humano [8], lo que se conoce con el nombre de *interacción natural*.

El desarrollo de este tipo de sistemas, denominados Sistemas de Interacción Natural, no es sino la lógica consecuencia del desarrollo tecnológico. A medida que los sistemas desarrollados ganan en complejidad y se extienden de forma irreversible a todos los aspectos de la vida de las personas, estos deben adaptarse también al entorno y a sus potenciales usuarios [10]. La tecnología debe hacerse transparente, integrarse en el mundo que las rodea y hacerse accesible por medio de una interacción lo más natural posible [188].

Los avances realizados en este tipo de paradigma interactivo resultan especialmente útiles para personas que, por acarrear algún tipo de discapacidad, vean restringidos los modos, códigos y procedimientos comunicativos en que pueden desarrollar la interacción. No se trata de personas imposibilitadas para la interacción, sino de personas con unas capacidades interactivas distintas. Cuales quiera que sean la formas de comunicación válidas para cada individuo, la interacción natural pretende hacerle accesible la tecnología adaptando los mecanismos interactivos aplicados a sus necesidades, pero sin limitar su propia capacidad interactiva. Estas limitaciones pueden ser de naturaleza cognitiva, visual, motriz, etc., y también se incluyen en

ellas la falta de entrenamiento para el uso de determinada tecnología (la brecha tecnológica) y cualquier otro tipo de dificultad que, tanto de forma crónica como temporal, pudiesen suponer una barrera para el acceso del usuario a la tecnología. De esta forma, cualquier persona es, en mayor o en menor medida, un potencial usuario con algún grado de discapacidad. Por tanto, un potencial beneficiario de la interacción natural.

Por ello se hace necesario afrontar la HCI desde un punto de vista radicalmente opuesto al tradicional: normalmente es el usuario quien debe adaptarse a las interfaces que le ofrecen los sistemas de interacción cuando el objetivo último de la interacción natural sería conseguir que todo el esfuerzo de adaptación recayese del lado de la máquina y no del lado humano. Conseguir que la tecnología sea capaz de desarrollar una interacción más natural a las personas se ha convertido en uno de los principales retos a alcanzar en la HCI, y los esfuerzos en esta línea no quedan circunscritos al ámbito académico, sino que son muchos los dominios de interacción comerciales en los que estas nuevas tecnologías están suponiendo importantes avances en la interacción con el usuario. Algunos ejemplos son:

- Atención al cliente
- Asistencia técnica
- Reservas de hoteles, vuelos o alquiler de vehículos
- Navegación y guiado
- Monitorización de pacientes y asistencia sanitaria
- Educación y tutores inteligentes
- Videojuegos

Entre los centenares de compañías que apuestan por el desarrollo de estas tecnologías se encuentran fabricantes de sistemas informáticos (IBM, Siemens AG), automóviles (BMW, Mercedes-Benz), operadores de telecomunicaciones (Telefónica I+D, Alcatel, France Télécom), compañias aéreas (Air France) o agencias aeroespaciales (Aerospatiale). Del mismo modo, son diversas las organizaciones que participan activamente en el desarrollo de estos sistemas, como el instituto Fraunhofer, el Centro Alemán de Investigación en Inteligencia Artificial (DFKI GmbH), la Agencia Estadounidense de Investigación en Proyectos Avanzados de Defensa (DARPA), o las principales universidades que investigan en el ámbito de las *tecnologías de información* (TI) o la HCI.

De esta forma, los Sistemas de Interacción Natural han experimentado un espectacular desarrollo durante los últimos años, especialmente en los aspectos relacionados con el reconocimiento y la síntesis de habla, y el tratamiento de la multimodalidad y el multilingüismo

en la interacción. En lo que respecta a la habilidad del sistema para estructurar y ordenar el conjunto de ideas desarrolladas en la interacción, entender el efecto que tiene sobre ellas lo que hacen o dicen los participantes y decidir qué debe hacer o decir el sistema a lo largo de la interacción (denominada gestión del diálogo), pueden distinguirse tres niveles distintos de organización [32]. Estos son: *organización global* (representación de las relaciones existentes entre las distintas metas que se desarrollan en la interacción); *local* (representación del desarrollo interno de cada una de las metas); y *temporal* (momentos en que los interlocutores deciden participar en la interacción y forma según la cual van desarrollando sus metas a lo largo del tiempo).

De estos tres niveles de organización, tanto la organización global como la local han experimentado importantes avances en los últimos años. Por su parte, la organización temporal (más comúnmente denominada *toma de turno*) ha sido, en la práctica, un aspecto poco tratado hasta el momento por los estudios en interacción natural. Por lo general, los Sistemas de Interacción Natural tienden a simplificar la toma de turno a un proceso de paso de testigo denominado *ciclo de interacción*. Según esta toma de turno, se considera que:

- Los participantes contribuyen en la interacción de uno por uno y en orden.
- El participante que contribuye decide unilateralmente durante cuánto tiempo lo hará y qué contenidos desarrollará en su contribución.
- Las contribuciones de los participantes no sufren ninguna modificación a lo largo del tiempo y, por tanto, son expresadas exactamente tal y como fueron formalizadas inicialmente, sin que exista ningún factor que pueda alterarlas dinámicamente mientras van siendo producidas.

Sin embargo, la toma de turno sí ha sido un proceso ampliamente estudiado para la interacción humana desde disciplinas como la lingüística, la sociología y la sociolingüística. Según éstas, la interacción que desarrollan las personas es una *acción combinada* [147] y, como tal, una acción colaborativa sobre la que todos los participantes tienen sus propios intereses y sobre la que intentan alcanzar un compromiso como medio para satisfacerlos. Cualquiera de las acciones que comprende la interacción es, en sí misma una acción combinada. Así lo es tanto la construcción de las contribuciones de los participantes, como la propia distribución de los turnos en la interacción. Según el *principio de la interpretación combinada* [77] todos los participantes de la interacción contribuyen a la coconstrucción de de las contribuciones del hablante. Esto es conseguido a partir de las muestras públicas de aceptación o rechazo de las acciones del hablante que los oyentes expresan a través de contribuciones de realimentación simultáneas, como lo son las contribuciones de asentimiento o rechazo, o expresiones como

"*ajam*" entre otras. Las muestras públicas son, habitualmente, expresadas a través de un flujo alternativo de acciones (*pista secundaria*) a las acciones con las que el hablante realiza su intervención (*pista primaria*). Por su parte, la distribución de turnos surge dinámicamente de las conjeturas que los participantes establecen sobre el compromiso común y los intereses particulares de cada participante, y puede verse alterada por los cambios ocurridos sobre las *circunstancias sociolingüísticas* en las que se desarrolla la interacción [149]. En definitiva, la distribución de turnos no está regulada por un ciclo rígido y predefinido de contribuciones de uno y otro participante, sino que constituye un proceso espontáneo e impredecible que se produce como consecuencia de la confluencia de los intereses particulares de los distintos participantes en un mismo compromiso común y que no puede ser entendida con independencia de las circunstancias en las que se produce [63, pp.39-46]. En la interacción humana las contribuciones de los participantes se van produciendo sobre la marcha y, en ella, fenómenos como los solapamientos, las interrupciones, así como otros tipos de disfluencia (rectificaciones, repeticiones, omisiones, silencios, silencios oralizados, etc.) [12], lejos de ser anomalías, resultan ser recursos frecuentes, necesarios y naturales.

Por todo ello, una interacción natural, desde el punto de vista de una toma de turnos avanzada, no puede restringirse a un proceso secuencial en el que cada uno de los problemas involucrados pueda ser abordado en un punto concreto de la ejecución [36]. Aunque para determinados dominios puede ser una aproximación válida (dominios de recuperación de información o transaccionales, en la mayoría de los casos), a medida que mejoran las habilidades interactivas de los sistemas de interacción (especialmente en la representación de las circunstancias sociolingüísticas de la interacción y en las capacidades pro activas de los Sistemas de Interacción Natural) son necesarios nuevos paradigmas para la gestión temporal de la interacción.

Durante la interacción natural son varios los procesos interactivos que deben desarrollarse simultáneamente, realimentándose unos a otros continuamente, para poder resolver los complejos problemas que ésta plantea. La interacción natural se desarrolla sobre los resultados parciales que los diferentes mecanismos de razonamiento van generando e intercambiándose entre ellos (simultáneamente a la producción de las contribuciones) y que van refinándose y realimentándose en tiempo real. Más allá de ser un proceso de iniciativa mixta (en el que la iniciativa de hacer progresar la interacción puede provenir de cualquiera de los participantes), es un proceso en el que la decisión de toma de turno de alguno de los participantes puede ocurrir en cualquier momento, incluso mientras otros participantes pudieran estar contribuyendo. Esta capacidad pro activa, natural a la interacción humana, requiere que los distintos componentes del sistema puedan evaluar en todo momento la necesidad de provocar o

no nuevos eventos en el sistema, independientemente del estado de la interacción y fuera de cualquier orden predefinido en la ejecución de los procesos involucrados.

Por todo ello, una interacción con una toma de turno avanzada requiere la aplicación de *nuevas arquitecturas de Sistemas de Interacción Natural*, que permitan *independizar los distintos procesos de la interacción natural* y con *capacidad para desarrollar estrategias avanzadas de presentación*.

1.1 *OBJETIVO*

El objetivo de esta tesis doctoral será analizar, definir, implementar y evaluar nuevos modelos de conocimiento para las arquitecturas de los Sistemas de Interacción Natural que hagan posible que este tipo de sistemas participen en interacciones desarrolladas bajo una *toma de turno avanzada*. Esta toma de turno será aquella en la que ni la distribución relativa de los turnos ni el número de participantes queden definidos de antemano, y en las que las situaciones de turnos solapados, interrupciones y disfluencias sean tratadas, no como fenómenos anómalos, sino como fenómenos frecuentes, aceptados y necesarios para el correcto desarrollo de la interacción. Las habilidades que deberá soportar el sistema para ello son:

La *adquisición e interpretación incremental* (sobre la marcha) de las contribuciones de los interlocutores. Esta habilidad consistirá desarrollar estos procesos simultáneamente a la producción de dichas contribuciones de los interlocutores, y no sólo tras su completa expresión. Entre otras aportaciones, esta habilidad se requiere para hacer posible la generación de realimentación simultánea, que permite que los oyentes participen en la elaboración de las contribuciones de sus interlocutores.

La *generación y síntesis incremental* de las contribuciones del sistema. Esta habilidad consistirá en revisar durante el proceso de generación la contribución que el sistema se encuentra expresando, pudiéndola adaptar sobre la marcha a los cambios acontecidos en las circunstancias que rodean a la interacción. Esto hará posible adaptar dinámicamente la respuesta del sistema en función de la realimentación simultánea ofrecida por sus oyentes y de los eventos ocurridos durante su expresión.

Una *decisión de toma de turno*. El sistema deberá ser capaz de identificar las situaciones en las que su participación en la interacción es adecuada y según qué tipo de contribución. Así mismo, esta decisión implicará llevar a cabo las siguientes acciones:

- Solicitar la palabra cuando el sistema requiera intervenir y las circunstancias no le sean propicias para ello.
- Tomar la palabra cuando el sistema requiera intervenir en la interacción y las circunstancias así lo permitan. Rechazarla cuando se espere de él que intervenga en la interacción pero éste no requiera hacerlo.
- Mantener la palabra cuando, aspirando otro participante a tomarla, el sistema considere que le corresponde mantenerla y las circunstancias lo permitan. Cederla en caso contrario.

Y, finalmente, *gestión de turnos*. Con el objetivo de dotar al sistema de elementos de juicio suficientes para desarrollar la decisión de toma de turno, deberá conjeturar el estado en el que se encuentran los turnos de los distintos participantes (así como el suyo propio), el estado de posesión de la palabra y qué participantes han sido designados o son candidatos a tomarla.

1.2 ESTRUCTURA DEL DOCUMENTO

El documento se apoya en estudios teóricos realizados desde la lingüística, sociolingüística y psicología sobre la forma en la que las personas usan del lenguaje en sus intercambios espontáneos [Capítulo 2]. A partir de ello, y tomando como punto de partida las tecnologías de Sistemas de Interacción Natural existentes hasta el momento [Capítulo 3], define y estructura una arquitectura cognitiva para Sistemas de Interacción Natural [Capítulo 4] en la que tienen cabida los nuevos modelos de conocimiento requeridos para el desarrollo de una toma de turno avanzada. Posteriormente se definen los objetivos específicos a alcanzar [Capítulo 5] y una propuesta de Sistema de Interacción Natural para una *toma de turno avanzada* [Capítulo 6]. Este trabajo de tesis doctoral también describe la evaluación de la arquitectura propuesta sobre el prototipo de sistema de interacción natural LaBDA-Interactor, de acuerdo a una metodología de evaluación especialmente diseñada para abordar los aspectos de toma de turno [Capítulo 7]. Finalmente, se describen las conclusiones alcanzadas con la realización de este trabajo y algunas de las líneas futuras de investigación que se desprenden de los resultados obtenidos [Capítulo 8].

Capítulo 2 FUNDAMENTOS TEÓRICOS

La organización temporal de la *interacción natural*, denominada *toma de turnos*, ha sido escasamente aplicada a los Sistemas de Interacción Natural hasta el momento. Tradicionalmente, el proceso interactivo ha sido modelado como el paso de un testigo, donde en cada momento la palabra está en posesión de un participante concreto (el *hablante*), y sólo cuando éste ha finalizado su participación pasa al siguiente, quien de nuevo podrá utilizarla cuanto tiempo desee antes de volver a pasarla. Desde este planteamiento, la participación del sistema es sólo la consecuencia de una participación previa del usuario y, del mismo modo, se espera que el usuario intervenga cuando el sistema termina de producir su intervención.

Si de lo que se trata es de reproducir algunas de las habilidades interactivas humanas, el punto de partida deberá ser el estudio de los trabajos de la lingüística, sociología y sociolingüística que tratan sobre la forma en que las personas interactúan entre ellas. Este problema ha sido ampliamente tratado en las últimas décadas desde numerosos enfoques, como son la *lingüística sistémico-funcional* [53], la *sociología de género* [167], la *visión co-construccionista* [118], la *sociolingüística interaccional* [83; 156] y la *etnometodológía* [149].

De todas estas líneas de investigación, la etnometodología, a través del *análisis conversacional*, trata de describir la estructura y los patrones que subyacen a cualquier tipo de conversación desarrollada entre personas (especialmente de aquellas que se producen de forma espontánea) y prestando atención, no tanto al lenguaje propiamente dicho, sino a la forma en la que este se utiliza.

Según esta corriente, la interacción natural no se trata de una sucesión de contribuciones independientes, producidas en un momento y orden concreto, tal y como postula el *ciclo de la interacción* [Apartado 2.1]. La interacción entre personas surge de forma imprevisible como

fruto de la colaboración interesada entre ellos. Se produce de forma oportunista, pieza por pieza, según los participantes van negociando sus *acciones combinadas*, a diferentes niveles, en diferentes planos y sobre diferentes capas [Apartado 2.2]. Considerar todos estos principios permite explicar de forma precisa la manera en la que los participantes estructuran la interacción, tanto a nivel *global* como *local* [Apartado 2.3], y permite determinar los principios que rigen la *producción de los turnos* y la *toma de la palabra* en la interacción [Apartado 2.4]. Por todo ello, el enfoque etnometodológico es aquel sobre el que se sustenta el presente trabajo.

2.1 *EL CICLO DE INTERACCIÓN*

El *ciclo de interacción* parte de la visión tradicional de la interacción que describe Saussure [151, pp.76-77], según la cual en toda comunicación hay un elemento activo, el emisor, y otro pasivo, el receptor. En consecuencia, el diálogo es una acción en la que cada participante desempeña uno de estos roles, bien el de emisor o el de receptor, y de ninguna forma puede tomar simultáneamente ambos. Bajo este planeamiento, la interacción es una acción dirigida desde el hablante hacia el oyente, sin ningún flujo de comunicación producido en sentido contrario. Junto a esto, se considera que la palabra pasa de un participante a otro a lo largo de la interacción, en un orden predefinido y fijo (a modo de testigo). El resultado es una interacción consistente en un proceso único en el que se suceden una y otra vez un conjunto de cinco fases [Figura 1]: *adquisición*, *interpretación*, *operación*, *generación* y *síntesis*. En parte del ciclo es el sistema quien toma el rol de oyente (adquisición e interpretación), y en parte el de hablante (generación y síntesis). En la fase intermedia (operación) aplica la tecnología a la que el usuario quiere acceder a través de la interacción.

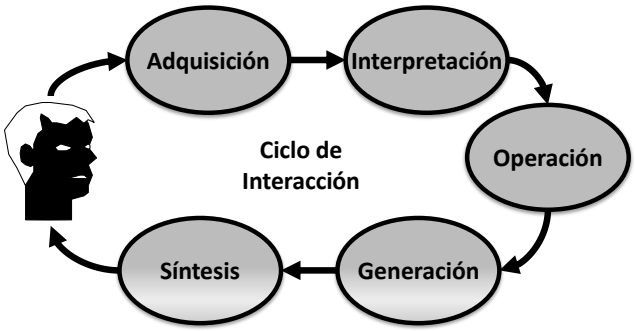


Figura 1: Ciclo de Interacción Tradicional en los Sistemas de Interacción Natural

Dicho desarrollo temporal es adecuado para algunos casos de interacción con los usuarios, especialmente cuando se trata de dominios transaccionales u orientados a la recuperación de información. El Ejemplo 1 muestra un caso particular de escenario en el que la

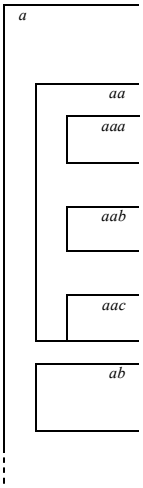
interacción sigue el mencionado ciclo. El ejemplo se estructura en los segmentos [154] *a*, *aa* y *ab*, cada uno de ellos compuesto por varias intervenciones (*a* por *a1*, *aa* por *aa1* y *aa2*, y *ab* por *ab1* y *ab2*). A pesar de mantener una correcta organización local y global (las relaciones entre las intervenciones de cada segmento y las relaciones entre los distintos segmentos, respectivamente), el tradicional ciclo de la interacción está caracterizado por una excesiva rigidez en cuanto a la toma de turnos y a la construcción de las intervenciones (en comparación con un intercambio real producido entre las personas en un uso real del lenguaje). De hecho, algunas cuestiones como el no hacer coparticipes a los oyentes de la producción de las intervenciones del hablante, o el no adaptar el contenido de los mensajes a la cantidad de información que requirieren, resultan prácticas antinaturales y, en muchos casos, agresivas en la *interacción humana*.

a	a_1	*U*	*¿Cómo puedo salir del edificio?*
aa	aa_1	*S*	*El edificio tiene dos salidas, la puerta principal y la de empleados. La más cercana es la principal, que está bajando las escaleras del final del pasillo. La otra está hacia atrás, en la zona B, junto al parque.*
	aa_2	*U*	*Vale, quiero salir por la del parque.*
ab	ab_1	*S*	*Entonces dese la vuelta y avance 10 metros, hasta la intersección con el siguiente pasillo.*
	ab_2	*U*	*Muy bien*
			...

Ejemplo 1: Interacción según el ciclo de interacción

Un escenario más cercano a una interacción entre personas sería el mostrado en el Ejemplo 2. En esta interacción, ni la duración, ni el contenido de las intervenciones queda definido unilateralmente por el hablante. Tal es el caso de la intervención compuesta por las contribuciones *aa1 + aa1' + aa1''*. Dicha intervención es coconstruida por ambos participantes a partir de una formalización previa del hablante (la intervención *aa1* del primer ejemplo) y reformulada a partir de la realimentación recibida del oyente (en las contribuciones *aaa1*, *aab1* y *aac1*). En ellas, *aaa1* ofrece evidencias de excesiva información. Esto lleva al sistema a reformular *aa1* en *aa1'*. Por su parte *aab1* expresa, a través de modalidades alternativas al habla, que en este caso hay carencia de información. Esto causa la reformulación de *aa1'* en *aa1''*. Por otro lado, la alteración en las circunstancias que rodean a la interacción, como se ve en el Ejemplo 3, también puede alterar el *ciclo de interacción* en la práctica. En este caso, la ocurrencia de eventos mientras se está desarrollando la intervención *ab1* fuerza al sistema a introducir la cancelación de la meta propia de recién abierto intercambio *ab*. Esto representa una auto interrupción del sistema.

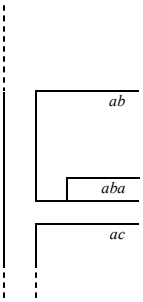
Al igual que ocurre en estos casos, es posible encontrar otros muchos ejemplos de situaciones en las que los cambios que se dan en la atención y el interés prestado sobre las metas desarrolladas, la evolución de la urgencia de determinadas metas, o su cancelación (entre otros factores) pueden tener efectos sobre el desarrollo temporal de la interacción.



a_1	*U*	*¿Cómo puedo salir del edificio?*
aa_1	*S*	*El edificio tiene dos salidas, la puerta principal y la de empleados. La más cercana es la principal, que está [bajando...]*
aaa_1	*U*	*[Ah, sí]. Ya sé [cual es.]*
aa_1'	*S*	*[Sí]. La otra salida es la de atrás*
aab_1	*U*	*(El usuario se muestra confuso)*
aa_1''	*S*	*...la de la zona B, la que está cerca del parque*
aac_1	*U*	*¡Ah, sí! Vayamos por ahí.*
ab_1	*S*	*Entonces dese la vuelta y avance 10 metros, hasta la intersección con el siguiente pasillo.*
ab_2	*U*	*Muy bien*
		...

Ejemplo 2: Interacción con coconstrucción de la intervención del sistema

En resumen, si de lo que se trata es de desarrollar una toma de turno natural, el enfoque del ciclo de la interacción resulta excesivamente rígido. Esto ocurre en aquellos casos en que se requiere un comportamiento más pro activo por parte del sistema o cuando la interacción es altamente dependiente de la alteración de las circunstancias sociolingüísticas en las que se desarrolla. Bajo estos supuestos, se requiere una toma de turnos más compleja, y ésta será fundamentada en aquellas teorías sociolingüísticas que describan de la mejor forma posible el modo en la que las personas organizan su interacción a lo largo del tiempo.



		...
ab_1	*S*	*Entonces dese la vuelta y avance...*
	E	*(El guardia de seguridad cierra la puerta)*
aba_1	*S*	*...Disculpe. El guardia de seguridad acaba de cerrar la puerta.*
ac_1		*¿No te importa si salimos por la otra?*
		...

Ejemplo 3: Interacción con auto interrupción del sistema

2.2 *TEORÍAS DE LA ACCIÓN COMBINADA*

La *interacción natural* es una actividad social que precisa de la acción conjunta de varios agentes. Esta actividad es desarrollada a través de acciones individuales que los participantes coordinan convirtiéndolas en *acciones combinada*s [147], y tal coordinación se da tanto en el *contenido* (aquello que los participantes intentan llevar a cabo) como en el *proceso* (los sistemas físicos y mentales a los que recurren para llevar a cabo sus intenciones).

El entender la interacción como una acción combinada es una idea inspirada en los estudios de Levinson [116; 117]. Para él, una *acción combinada* es una categoría borrosa que representa una meta acotada y constituida socialmente, y que describe restricciones dadas por los participantes y el entorno. Herbert H. Clark [32], basándose en los trabajos de Levinson, clasifica cualquier tipo de acción combinada según cinco rasgos distintos:

- *Guionizadas / Improvisadas*: Nivel en el que la acción sigue un guión o protocolo previamente establecido. Acciones altamente preparadas son, por ejemplo, las ceremonias nupciales y en el lado opuesto se encontraría un encuentro casual en la calle.
- *Formales / Informales*: En función de los niveles de exactitud, puntualidad, seriedad y consecuencia manifestados en la acción. En el extremo formal se encontraría, por ejemplo, un discurso político. Ejemplo de acción informal sería comentar un chismorreo.
- *Verbales / No verbales*: Proporción de la acción que se desarrolla de forma hablada o por otros medios. Llamada telefónica frente a un partido de fútbol.
- *Cooperativas / Competitivas*: En función de peso que tengan en la acción alcanzar objetivos comunes (por ejemplo, hacer negocios) o particulares (ganar un partido).
- *Autocráticas / Igualitarias*: Nivel en que la acción está gobernada o liderada por uno de los participantes. Una acción muy autocrática sería una lección, mientras que el hecho de conocerse distintos individuos sería una acción altamente igualitaria.

A partir de la definición de acción combinada y de sus propios estudios, Herbert H. Clark identifica los distintos elementos que comprende la acción combinada:

- *Participantes,* desempeñando determinados roles.

- *Metas* (compartidas o individuales), las cuales los participantes tratarán de alcanzar por el bien común o en busca de su propio beneficio.

Para que una acción combinada pueda desarrollarse, los participantes deben hacer públicos sus roles en la acción. Los cometidos que asuman en la acción y su responsabilidad dependerán del rol que desempeñen. A través de una misma acción combinada los participantes pueden perseguir más de una meta [17; 70; 90], algunas pueden ser procedimentales (como desarrollar la acción tan rápido y eficientemente como sea posible), otras interpersonales (ser educado y respetuoso, entre otras) y privadas (por ejemplo, aprovechar la situación para obtener ventajas personales). En cualquier caso, en función de su influencia en el desarrollo de la acción combinada pueden ser de dos tipos:

- *Metas compartidas*: Aquellas metas que son reconocidas por todos los participantes y pueden ser establecidas explícitamente (decidir jugar al parchís) o implícitamente (el tendero da por supuesto que el cliente desea comprar lo que lleva en el carro).
- *Metas privadas*: El resto de metas que se desarrollan tras la acción. En general son metas desarrolladas en beneficio propio. Frecuentemente las metas privadas entran en conflicto con alguna de las metas compartidas desarrolladas en la actividad, pero los participantes deben desarrollar las metas compartidas si buscan el éxito de la acción completa y, con ello, alcanzar el resto de sus metas individuales.

Cualquier meta compartida debe ser iniciada por alguno de los participantes y acordada por el resto. Por tanto, para el desarrollo de la acción combinada, es necesario que los agentes acepten las metas compartidas. La aceptación del las metas compartidas implica el conocimiento de las mismas y también el compromiso a desarrollarlas. El cumplimiento de estas dos condiciones es lo que permite la coherencia en la interacción.

Por "conocimiento de las mismas" entendemos "conocimiento mutuo de las mismas", definiéndose el conocimiento mutuo de los participantes (A y B) acerca de un hecho X como:

1. A conoce X.
2. B conoce X.
3. A y B conocen 1, 2 y 3.

Por otro lado, un participante aceptará el compromiso a desarrollar las metas compartidas siempre que crea que el resto de participantes lo ha asumido también. El compromiso se mantendrá mientras alguno de los participantes considere que la meta no ha sido alcanzada. No obstante, puede ocurrir que la conjetura individual de uno de los participantes sobre el compromiso de una meta no se ajuste a la realidad, debilitando el conocimiento mutuo sobre la misma.

Las acciones combinadas pueden estar compuestas de acciones menores, muchas de las cuales pueden ser también acciones combinadas. También deben estar caracterizadas por un principio, un desarrollo (opcional) y un fin para alcanzar el éxito. Pueden ser simultáneas o intermitentes, y se pueden expandir, contraer o dividir en otras menores.

Cada uno de los participantes conoce en todo momento sus propias metas, pero para poder desarrollar la interacción, también debe estimar el conjunto de metas compartidas que se están desarrollando, así como el estado en el que supone que se encuentra cada una de ellas. Esta representación del la interacción como una acción combinada es lo que se denomina *zona común* [Apartado 2.2.1]. Por otro lado, el hecho de que la interacción se desarrolle en forma de acción combinada implica que las acciones que la componen no podrán ser realizadas de forma unilateral por ningún participante. Por ello, toda acción desarrollada en la interacción requerirá otra acción complementaria (*evidencia de cierre*) producida por el resto de participantes [Apartado 2.2.2]. Finalmente, podrán caracterizarse las distintas dimensiones (*líneas de acción*) en las que actúan las acciones realizadas en la interacción [Apartado 2.2.3].

2.2.1 La Zona Común

El desarrollo de las acciones combinadas se produce paso a paso de forma incremental. Cada participante posee una *zona común* [32, pp.92-124] donde va acumulando el conocimiento, creencias y suposiciones que cree compartir con el resto de participantes sobre las actividades combinadas que se desarrollan en la interacción y que va construyendo a lo largo del tiempo. La posesión de una zona común por parte de cada participante es lo que hace posible que la interacción progrese. Según Stalnaker[163], la zona común está compuesta por:

- *Zona común inicial*: Son el conjunto de hechos, presuposiciones y creencias de partida del participante en el momento de entrar a participar en cada acción combinada.
- *estado actual de la zona común*: La estimación del estado del desarrollo de cada acción combinada en cada momento.

- *Movimientos combinados realizados hasta el momento*: Es el conjunto de movimientos compartidos que el participante estima que se han realizado hasta llegar al estado actual que supone.

Dado que la zona común está compuesta por las presuposiciones propias de cada participante, a menudo existen discrepancias entre la acción combinada que se desarrolla y la forma en la que la refleja la zona común de algún participante. Es decir, puede ocurrir que las conjeturas individuales de los participantes sobre el compromiso de una meta discrepen [163, pp.321], debilitando el conocimiento mutuo sobre la misma.

2.2.2 Evidencias de Cierre

Desde el punto de vista clásico, la interacción surge del intercambio de acciones unilaterales entre los participantes, donde el hablante fija sus intenciones y los oyentes se limitan a identificarlas. En realidad la interacción es una acción combinada y, como tal, no consiste en acciones unilaterales desarrolladas por los participantes, sino de acciones acordadas y comprometidas conjuntamente [32, pp.221-252].

Alcanzar la interpretación de una señal no es sencillo. A ello contribuyen la forma de la enunciación y las creencias mutuas que los participantes tienen sobre las circunstancias en las que se desarrolla la interacción. Además de éstas, hay una fuente de información adicional que tradicionalmente ha sido omitida: las muestras públicas de las interpretaciones propias producidas por los oyentes durante la conversación y la aceptación o rechazo de estas muestras públicas por el hablante [131].

La idea principal es que hablantes y oyentes tratan de crear una interpretación combinada de lo que se supone que el hablante quiere decir. Tal interpretación representa, no lo que el hablante quiere decir en sí mismo (lo que puede cambiar durante el proceso de comunicación), sino lo que los participantes suponen mutuamente que el hablante quiere decir [78]. La idea se refleja en el siguiente principio:

Principio de la interpretación combinada: Para cada señal, el hablante y sus oyentes tratan de crear una interpretación combinada de lo se supone que el hablante que quiere decir con ella.

Con este principio el oyente no está simplemente tratando de identificar lo que el hablante quiere decir, sino que está intentando crear una interpretación que ambos estén dispuestos a aceptar como lo que quería decir el hablante. El oyente, normalmente, tratará de

inferir la intención inicial del hablante, pero la intención combinada a la que ellos llegan a menudo difiere de las intenciones originales del hablante. En realidad, para muchas señales, la idea clásica de "lo que el hablante quiere decir" no tiene sentido, pero sí la idea de "lo que se supone que el hablante quiere decir". Este principio apoya la idea que la interacción es un proceso en el que en todo momento todos los participantes toman parte, y no una acción unilateral desempeñada desde el hablante hacia los oyentes, tal y como se postula en el *ciclo de interacción*.

Los participantes asumen esta característica de la *interacción natural* y tratan de ofrecer al hablante evidencias de su interpretación, de forma que éste pueda validarla o corregirla. La primera alternativa (y la más comúnmente usada) consiste en mostrar al hablante la aceptación correspondiente a la interpretación realizada, dándola como buena [33; 34]. De esta forma pueden distinguirse entre varias posibles aceptaciones para cualquier primera parte de par propuesta, cada una de ellas correspondiente a cada una de las posibles interpretaciones posibles. Así por ejemplo, ante la enunciación "*Siéntate*" por parte del hablante, las alternativas que tiene el oyente utilizando esta estrategia de evidencias de cierre son las mostradas en la Tabla 1.

Tabla 1: Posibles aceptaciones de la propuesta "*siéntate*"

Enunciación del hablante	Interpretación del oyente	Enunciación de Aceptación
Siéntate	Una orden	*Sí, gracias*
Siéntate	Una respuesta	*Vale*
Siéntate	Una oferta	*No, gracias*
Siéntate	Un consejo	*Qué buena idea*

Al responder el oyente con una aceptación, muestra al hablante la interpretación que ha realizado. A partir de este momento depende del hablante llegar a una interpretación combinada. Cuando el hablante acepta la interpretación ésta pasa a ser la *interpretación combinada*. Si, por el contrario, la interpretación evidenciada por el oyente no encaja con la del hablante, éste tendrá la oportunidad de corregirla para refinar dicha interpretación combinada:

Hablante – Siéntate
Oyente – No, gracias
H – Es una orden

Al considerar este principio, el oyente pasa de ser un agente pasivo (que simplemente identificaba lo que el hablante quería decir) a ser una parte activa en la producción de las intervenciones del hablante, capaz de elaborar suposiciones sobre lo que pretende decir y de mostrarlas para que el hablante las acepte o rechace. El oyente, por tanto, colabora en la construcción de las intervenciones del hablante, a través de sus contribuciones que, a menudo,

son realizadas de forma simultánea a la del propio hablante. Con ellas afecta a la formalización y generación de la intervención que el hablante se encuentra produciendo.

En definitiva, la interacción es un proceso en el que todos los participantes toman parte en todo momento y, con ellos, pasa de considerarse una sucesión de ciclos secuenciales a ser la convergencia de múltiples acciones desarrolladas simultáneamente por todos los participantes.

2.2.3 Líneas de Acción

La interacción es construida a través de las acciones individuales que los participantes comprometen, convirtiéndolas en acciones combinadas. Cada una de estas acciones podrá ser caracterizada en un espacio de dimensiones denominadas *líneas de acción*. Estas dimensiones son *niveles*, *pistas* y *capas de acción* [32, pp.388-391].

La dimensión básica para producir la interacción son los niveles de acción, pero las pistas y capas son las que hacen posible la metacomunicación y la representación de situaciones ficticias dentro de la propia conversación.

2.2.3.1 Niveles de Acción

En el análisis convencional, las acciones combinadas se organizan en una jerarquía de cuatro niveles distintos, donde cada uno de ellos es, por si mismo, una acción combinada. Estos son *nivel de comportamiento*; *de señal*; *de mensaje*; y *nivel de proyecto combinado*:

- *Nivel de comportamiento*. Siendo *A* el emisor y *B* el receptor, la acción de *A* consiste en ejecutar un comportamiento *c* para *B* y la de *B* en percibir *c* de *A*. Este nivel requiere la atención de *B* a la voz y gestos del participante *A*.
- *Nivel de señal*: El comportamiento *c*, ejecutado por *A* y recibido por *B*, contiene una señal *s* de *A* para *B*. En este nivel se requiere de *A* que presente la señal *s* para *B* y de *B* que la identifique. Esto requiere que *B* identifique el conjunto de símbolos que comprende la señal de *A*.
- *Nivel de mensaje*. En este nivel lo importante es, por un lado, lo que *A* pretende decir y, por otro, lo que *B* entiende con el mensaje *m* que contiene la señal *s*.
- *Nivel de proyecto combinado*. Tras el mensaje *m* se esconde una proposición de proyecto combinado *p* de *A* hacia *B*. *B* deberá considerar *p* para comprometerse con él o rechazarlo.

Por ser las acciones de cualquier nivel acciones combinadas, cualquiera de ellas surge de la combinación de las acciones independientes y separadas del hablante y de los oyentes. Así

mismo, en todas ellas se requieren las evidencias de cierre que permiten desarrollar una interpretación combinada. Los niveles de acción tienen la propiedad de extender hacia arriba causalidad (la identificación de un proyecto parte de la interpretación de un mensaje, el mensaje de la identificación de una señal, y la señal de la detección de un comportamiento) y llevar hacia abajo la evidencia (la aceptación de un proyecto implica la correcta interpretación del mensaje en que se proponía, y la correcta interpretación de este mensaje evidencia la identificación de la señal que lo portaba, y ésta la detección del comportamiento con que se ejecutaba).

2.2.3.2 Pistas de Acción

Gestionar la interacción es una actividad distinta su propio desarrollo. Por esta razón, en toda interacción pueden distinguirse dos flujos distintos de acciones:

- *Flujo primario*: Representa el propio desarrollo de los "asuntos oficiales" de la interacción (que suele incluir el habla fluida y los intercambios suaves). Es todo aquello de lo que realmente trata la interacción.
- *Flujo secundario*: Conjunto de intercambios que tratan de crear una comunicación exitosa y se utilizan para su gestión (mostrar interés, solicitar aclaraciones, gestionar la toma del turno, etc.). Las contribuciones secundarias más comunes se realizan a través del *paralenguaje* [139; 142], aunque también pueden ser verbales (si el tono y el volumen en el que se emiten no suponen una interferencia en el flujo primario).

Hablamos, respectivamente, de la *pista primaria* de la interacción y de su *pista secundaria* (o comunicación colateral). Tradicionalmente, se ha prestado una mayor atención a la comunicación primaria que a la secundaria, por tratarse ésta última de un fenómeno de baja frecuencia en las bases de estudio más comunes del lenguaje (textos y las citas). Sin embargo la interacción está más ligada al *uso del lenguaje* y a su producción en tiempo real. Es en estas circunstancias en las que la comunicación colateral toma un mayor protagonismo.

Las señales colaterales tienen que ver con el significado y el entendimiento. Así, por ejemplo, cuando se asiente con la cabeza, se sonríe o se expresa un "*ajam*" durante la intervención de otro, se le está diciendo "entiendo lo que has dicho hasta ahora", con lo que se le permite obtener un cierre a sus acciones primarias. La comunicación colateral puede utilizarse también para evidenciar problemas (mostrar extrañeza, expresar "*¿eh?*"), que podrían estar relacionados con el nivel de comportamiento, nivel de señal (problemas en el canal), de mensaje (problemas de interpretación), o de proyecto combinado (problemas en la aceptación del

proyecto). Del mismo modo, puede ser aplicada por el hablante para desarrollar rectificaciones ("*son las siete y cuart... perdón, las seis y cuarto*").

En general, cualquier contribución colateral puede encajar en alguno de los casos mostrados en la Tabla 2.

Tabla 2: Posibles significados e interpretaciones de las contribuciones colaterales

	Sobre el significado del hablante	**Sobre lo que entiende el oyente**
Hablante	Con x quiero decir y. ¿qué entendiste tú por x?	¿Entendiste y por x?
Oyente	¿Con x quieres decir y? ¿Con x que quieres decir?	Yo entiendo x Por x yo entendí y

Además, la comunicación colateral (por ser secundaria a la comunicación principal) debe ajustarse a un conjunto de normas:

1. Realimentación: Mientras que las señales sobre los asuntos oficiales de la interacción deben ser marcadas, las señales de la pista secundaria deben ser señales de fondo.

2. Simultaneidad: Si los participantes realizan acciones en ambas pistas, es preferible que las realicen simultáneamente. Esto puede suceder de cuatro formas distintas:

    a. El hablante puede desarrollar un mismo comportamiento a través de señales en ambas pistas.

    b. El hablante puede desarrollar una señal en la pista secundaria simultáneamente al desarrollo de otra en la pista primaria, pero de ser así debe hacerlo a través de modalidades diferentes (por ejemplo, una a través de habla y otra a través de gestos).

    c. El oyente puede desarrollar señales en la pista secundaria al mismo tiempo que el hablante está desarrollando señales en la pista primaria, en la misma modalidad o en otra diferente.

    d. El hablante puede desarrollar señales de pista secundaria durante los intersticios (pausas mínimas) de la pista primaria.

3. Brevedad: La mayoría de las señales colaterales llevan una cantidad de información suficientemente pequeña como para que puedan ser realizadas en tiempos reducidos.

4. Diferenciación: Las señales en la pista colateral necesitan distinguirse claramente de las señales de la pista primaria. Para ello se aplican los siguientes mecanismos:

    a. Colocación temporal: Los hablantes pueden indicar si ellos entienden o no lo que fue dicho por la colocación de sus enunciaciones [Tabla 3].

    b. Prosodia marcada: Cualquier enunciación en el canal 1 tiene una prosodia esperada. Una forma de crear una señal colateral es superponiendo una prosodia no esperada o *marcada* en la enunciación.

    c. Gestos: Los gestos normalmente son perfectos para las señales colaterales. Son fáciles de distinguir del habla y pueden desarrollarse simultáneamente, de forma breve, y de fondo. Movimientos de cabeza, por ejemplo, son utilizados como asentimientos. Algunos ejemplos de gestos con funciones metacomunicativas son mostrados en la Tabla 4.

Además, la distribución en pistas de la interacción admite una organización recursiva de las mismas, pudiendo la pista secundaria ser a su vez pista primaria a otra nueva pista ("*No te oigo*" / "*¿Qué no qué?*")

Tabla 3: Importancia de la colocación de contribuciones colaterales

Señal de la pista secundaria	Ejemplo	Interpretación
Asentimiento	*A:Ha sido un día muy bonito* *B:sí*	Entiendo que acabas de terminar
Marcador de incertidumbre	*A: Vale, el siguiente es el conejo* *B: ¿Eh?*	No he entendido aún lo que acabas de terminar
Terminación colaborativa	*A: Es que se me ha olvidado la cartera* *B: Vale, no tienes dinero*	Tú quieres decir esto: "*no tengo dinero*"
Truncamiento	*A: Dónde está el otro…* *B: Detrás de ti*	Ya te he entendido la pregunta, así que te voy a responder

De esta forma puede distinguirse entre el flujo primario de acciones (que representan un intento por llevar a cabo asuntos oficiales) y un flujo colateral o secundario a éstas (que trata de crear una comunicación exitosa). La principal utilidad de la comunicación colateral radica en que permite obtener continuas evidencias de lo que los oyentes están entendiendo (sin necesidad de que tomen la palabra), dando la posibilidad al hablante de ajustar su intervención

continuamente, frase por frase, al dar cuenta de las evidencias de cierre de sus acciones. No obstante, la comunicación colateral puede desempeñar otras funciones, como son la rectificación por parte del hablante sobre lo anteriormente expuesto, las solicitudes y cesiones de palabra, etc. Las contribuciones colaterales pueden ser expresadas a través de modalidades diferentes a la de la presentación primaria (gestos, movimientos de cabeza, etc.) o a través de esta misma modalidad (a menudo el habla). Así mismo, cualquier acción colateral podrá tener también acciones colaterales propias, encaminadas a favorecer su propio éxito.

Tabla 4: Ejemplos de gestos con funciones metacomunicativas

Tipo de gesto	Función del gesto	Ejemplo
Gesto de reparto	Para referirse al reparto de información del hablante para los oyentes	El hablante revela a un interlocutor nueva información relevante para el proyecto
Gesto de cita	Para referirse a una contribución previa de los interlocutores	El hablante apunta al interlocutor para indicar "como dijiste antes"
Gesto de solicitud	Para sonsacar una respuesta específica de un interlocutor	El hablante mira a los interlocutores como si dijera "*¿podéis decirme la palabra para…?*"
Gesto de turno	Para tratar cuestiones acerca del turno de habla	El hablante libera el turno para un interlocutor

2.2.3.3 Capas de Acción

A menudo la gente parece estar diciendo algo cuando realmente de lo que habla es de otra cosa. Austin [5] denomina a este fenómeno usos *no serios* de la enunciación, y es la línea acción del lenguaje en la que se basan novelas, obras de teatro, películas, historias, e incluso bromas, ironía, sarcasmo, tomaduras de pelo, exageraciones, subestimaciones, etc. En todas estas acciones es común el fenómeno denominado por H. Clark *organización en capas*.

Estas capas, que admiten también una organización recursiva, son las que permiten desarrollar situaciones que contienen una componente de ficción, así como los usos figurados (ironía, sarcasmo, hipérbolas, bromas, retórica, etc.) y aparentes del lenguaje (situaciones en las que el hablante, por ejemplo, realiza una invitación a su interlocutor sabiendo de antemano que será declinada). En definitiva, consiste en un recurso utilizado con gran frecuencia y que impregna la comunicación de un carácter y sentido imaginativo del que sin él carecería [181; 182; 19].

2.3 *ORGANIZACIÓN DE LA INTERACCIÓN*

Existe una organización que estructura la interacción, no como un conjunto de turnos aislados, sino como una serie de turnos donde cada uno está conectado con el resto de alguna

forma. El origen de esta organización entre turnos radica en que todos ellos están al servicio del desarrollo de los proyectos combinados que motivan la interacción [32, pp.191-220]. Estos proyectos se estructuran de forma jerárquica en forma de proyectos y subproyectos, y a su vez cada proyecto se desarrolla como secuencias de intercambios materializados a lo largo del tiempo en los turnos de los participantes. En definitiva, la interacción se puede estructurar en los siguientes niveles de organización [32, pp.319-352]:

- *Organización global*: Hace referencia a como se estructuran los intercambios desarrollados en la interacción. Surge como consecuencia de la expansión de unos intercambios con otros al principio, durante o al final de su desarrollo.
- *Organización local*: Queda determinada por las relaciones de relevancia y preferencia existentes entre los turnos pertenecientes a un mismo intercambio. El desarrollo de las primeras partes de un intercambio hace relevante el desarrollo de alguna de las posibles segundas partes, de las cuales una será preferente frente a las otras.
- *Organización temporal*: Esta organización determina la forma en la que se coordinan la producción de los turnos de los distintos hablantes en el tiempo y a las reglas que rigen la posesión de la palabra en la interacción.

Aunque todos estos niveles de organización deben ser considerados para gestionar de forma adecuada la toma de la palabra y la alternancia de los turnos en la interacción, se prestará especial atención a la organización temporal, de la cual depende el tratamiento de la toma de turnos en la *interacción natural*.

2.4 ORGANIZACIÓN TEMPORAL DE LA INTERACCIÓN

Aunque en la aplicación práctica de los Sistemas de Interacción Natural se tiende a considerar que la conversación se desarrolla siguiendo un protocolo secuencial [Apartado 2.1], son muchas las evidencias que muestran que el proceso de la toma de turno es un proceso más complejo, donde deben considerarse simultaneidad, concurrencia, interrupciones, solapamientos, realimentación, generación e interpretación incremental y disfluencias, y así lo muestran importantes estudios teóricos [72; 86; 111; 134; 155; 49; 88; 31; 73; 50; 85; 44; 149; 12].

En general, todos estos estudios ponen de manifiesto que los participantes colaboran con el objetivo de producir sus intervenciones en ausencia de interferencias serias. La mayor

parte de las interferencias que podrían atentar contra la salud de la interacción provienen de otros participantes y consisten en enunciaciones solapadas a su intervención. Por esta razón, los hablantes deberían tender a respetar un conjunto de acuerdos implícitos que favorecen la producción de intervenciones en condiciones de exclusividad sobre el canal. Lo ideal sería hablar de uno en uno, según *la regla del hablante único*, pero existen razones para que esta regla no sea seguida en todos los casos.

En primer lugar, el tiempo de los participantes es, en sí mismo, un recurso demasiado valioso como para tolerar grandes pausas entre las intervenciones de los distintos participantes. De esta forma los aspirantes a tomar la palabra tratarán de proyectar el final de la intervención del hablante en curso para predecir con precisión en qué momento exacto deberán empezar a intervenir. Esto, a menudo deriva en solapamientos entre el final de la intervención previa y el inicio de la nueva, por problemas de temporización o tras haber sido proyectado el final de la intervención y no ser necesario esperar a su final. Este tipo de solapamiento ocurre en los denominados *Lugares de Transición Pertinentes* (TRP, del inglés *Transition Relevant Places*) del turno[154, pp.3-12], y no es interpretado por los interlocutores como interferencias serias, sino más bien como todo lo contrario.

Además, el momento en el que se producen los turnos y el orden en el que se toma la palabra no quedan completamente determinados por la regla del hablante único, por lo que se requieren criterios adicionales que permitan establecer el orden de preferencia entre los posibles candidatos a tomar la palabra (minimizando la colisión de las intervenciones de los distintos participantes) y seleccionar algún candidato cuando no los haya (evitando dejar el canal inactivo). De todos los modelos propuestos para representar el traspaso de la palabra en la *interacción natural*, el más relevante es el propuesto por Sacks et al. [149]. Según éste modelo, el hablante que está en posesión de palabra queda determinado de la siguiente forma:

1. Para cada TRP de la intervención en curso:

    a. En el caso de que el turno actual esté construido de tal forma que implique la selección del hablante siguiente (por ejemplo por haberle señalado con una acción que requiera una repuesta como una disculpa, una excusa, etc.), el participante seleccionado tendrá el derecho y la obligación de tomar la palabra. Nadie más tiene ese derecho ni esa obligación, y el traspaso ocurrirá en ese momento.

    b. Si el turno actual no está construido de tal forma que implique la selección del hablante siguiente, cualquier participante podría elegirse a

sí mismo como el hablante siguiente, aunque no de forma obligatoria. El primero en comenzar su intervención adquiere el derecho sobre la palabra y la transferencia ocurrirá en ese momento.

c. Cuando el turno actual no implica la selección del hablante siguiente y ningún otro participante se ha apropiado de la palabra, el hablante actual podría continuar hablando, aunque no de forma obligatoria.

2. La regla 1 será aplicada sucesivamente hasta haber sido realizado el traspaso de la palabra.

Aunque sólo de la simple aplicación de estas reglas se deduce que ni el orden ni la duración de los turnos está definido a priori, existen otras razones adicionales que hacen de la toma de turno de la interacción humana un proceso difícilmente conciliable con el denominado *ciclo de la interacción*. La principal de todas ellas es que, en la práctica, ni regla del hablante único ni el modelo de traspaso de la palabra son normas de obligado cumplimiento. Sólo definen el conjunto de acuerdos tácitos que, en ausencia de otros factores, las personas parecen respetar durante la interacción. En la práctica, aunque no es recomendable violar dicha regla, existen numerosas excepciones que son plenamente aceptadas por los participantes y que derivan en la ocurrencia de turnos solapados e interrupciones. Las principales son [32, pp.323-329]:

- Las contribuciones con fines metacomunicativos (las que se realizan sobre el plano secundario), que pueden ser desarrolladas sin la posesión de la palabra (aun solapando con la presentación primaria en curso).
- Las intervenciones justificadas por las circunstancias sociolingüísticas en las que se desarrolla la interacción. Entre ellas se encuentran la dominancia entre los roles de los participantes o las preferencias de unos usuarios frente a otros.
- Las intervenciones primarias que pueden ser presentadas a través de modalidades alternativas a la de la presentación primaria en curso.
- Por último, también es aceptado el solapamiento del principio de una nueva intervención con el final de la presentación el curso, con el doble objetivo de eliminar las pausas entre las intervenciones y de adelantarse a otros posibles candidatos a tomar la palabra.

En definitiva, la toma de turnos es un proceso no marcado, en el que ni el orden ni la duración de los turnos queda definido a priori y en el que, a pesar de ser preferible respetar unas reglas de traspaso de la palabra, existen multitud de situaciones en las que el solapamiento de

los turnos o las interrupciones no sólo no suponen una violación de las reglas de toma de la palabra, sino que resultan ser un recurso especialmente valioso para desarrollar con naturalidad la interacción.

2.5 *TIPOLOGÍA DE TURNOS*

Del conjunto de reglas descritas sobre la *organización temporal de la interacción* [Apartado 2.4], y considerando la clasificación de los posibles tipos de contribución de la interacción que propone Gallardo Pauls [63, pp.46-50], se estructura la tipología de turnos de la interacción como sigue [Figura 2]:

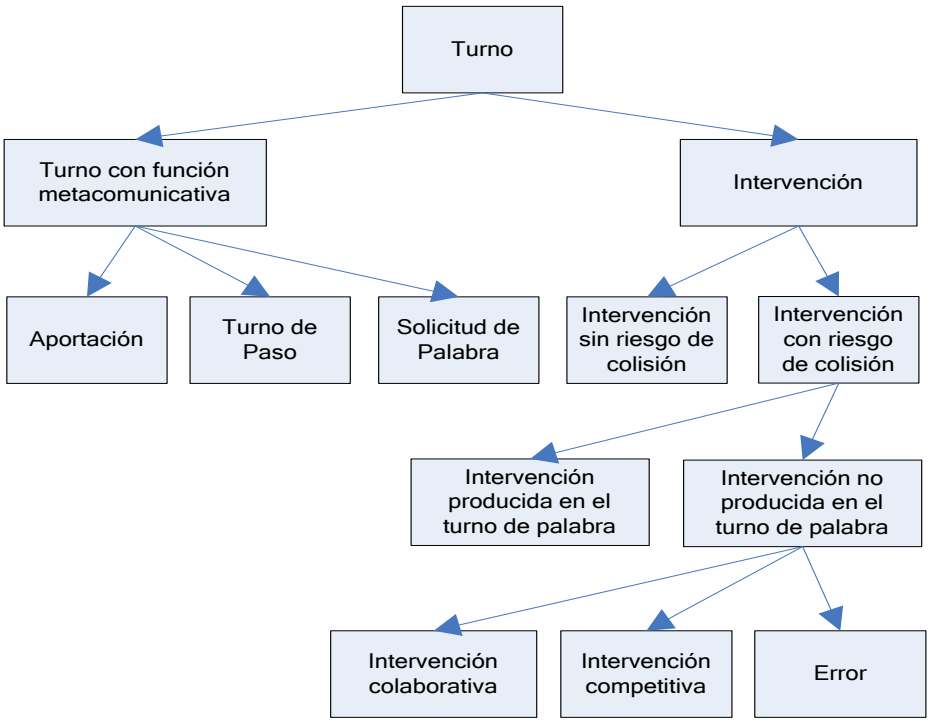


Figura 2: Taxonomía de turnos de la *interacción natural*.

1 *Turnos con funciones metacomunicativas*: Son presentaciones realizadas por completo en el plano secundario, con funciones de control sobre la propia interacción. Pueden ser *aportaciones*, *turnos de paso* y *solicitudes de palabra*.

1.1 *Aportaciones*: Son los turnos que constituyen la comunicación de realimentación que ofrecen los oyentes al hablante durante su presentación primaria y que se producen, por lo general, en situaciones de solapamiento (sin constituir en ningún caso violación de la

toma de la palabra). Este tipo de turnos pueden ser continuadores (que ratifican la distribución de papeles), reconocedores (que dan la razón o se la quitan al hablante), muestras de interés, peticiones de aclaración, ayuda a la búsqueda de expresiones o palabras y, en algunos casos, la risa.

1.2 *Solicitudes de palabra*: Las solicitudes de palabra son aquellas presentaciones con las que un determinado participante expresa su pretensión de intervenir y ocurren durante la intervención de otro participante.

1.3 *Turnos de paso*: Son turnos producidos por el hablante cuando recae sobre él la palabra de forma natural, pero desea rechazarla. Ocurren sobre todo en secuencias de cierre.

2 *Intervenciones*: Las intervenciones son los turnos que desarrollan los asuntos oficiales de la interacción (plano primario), aunque pueden incluir también expresiones metacomunicativas o de plano secundario. Cabe distinguir entre las que pueden suponer colisión con la intervención en curso y las que no.

2.1 *Intervenciones sin riesgo de colisión*: Casos en los que la intervención puede ser presentada sin suponer una colisión con la presentación previa, al ser producidas a través de modalidades no dominantes en la intervención previa, o incluso distintas.

2.2 *Intervenciones que pueden suponer riesgo de colisión*: No pueden ser presentadas a través de modalidades alternativas que no supongan una colisión en el canal con la presentación en curso. Entre ellas están las que se producen con la posesión de la palabra y las que no.

2.2.1 *Intervenciones producidas en el turno de palabra*: Son aquellas intervenciones producidas por aquel participante que, en un determinado momento, es reconocido por todos los participantes como el hablante, según las reglas que determinan la posesión de la palabra.

2.2.2 *Intervenciones no producidas en el turno de palabra*: En determinadas situaciones, un participante podría requerir tomar la palabra aún cuando no le corresponde. Aunque todas ellas constituyen violaciones de las reglas del traspaso de la palabra, no siempre son entendidas como tal por los participantes [60; 166]. Para evaluar correctamente este tipo de intervenciones, deben ser considerados la situación, los participantes, las metas emocionales, la propia interacción desarrollada, etc. [198]. Se distingue entre errores, intervenciones cooperativas e intervenciones competitivas.

2.2.2.1 *Errores*: ocurren cerca de un posible punto de transición de hablante, cuando un aspirante a tomar la palabra se anticipa y comienza a intervenir por error, al interpretar que la intervención en curso iba a finalizar aun sin ser esto cierto.

2.2.2.2 *Intervenciones colaborativas*: En estos casos, otros participantes se unen al hablante a concluir simultáneamente con él (y con sus mismas palabras) la intervención que está presentando. En algunos casos surge de la necesidad del oyente de participar en el discurso sin esperar a otro momento posterior, en el que tal vez su participación ya no sería tan pertinente, o con la intención de expresar entusiasmo compartido.

2.2.2.3 *Intervenciones competitivas*: Varios participantes intentan imponerse en el dominio de la palabra. En estos casos alguno de los hablantes tiene prioridad para intervenir sobre el resto (por el rol, por la situación, o por los propios usuarios). También pueden darse por haber estimado alguno de los participantes que, de no interrumpir, determinadas metas que pretende desarrollar podrían quedar olvidadas (cambio de foco o inserción de iniciativas) o con el objetivo de cancelar subdiálogos.

2.6 *CONCLUSIONES*

En la *interacción natural* el hablante en curso, aun en posesión de la palabra, no está en disposición de determinar durante cuánto tiempo la poseerá ni el contenido final de su intervención [Apartado 2.4]. El hablante podrá tener una formalización previa de aquello que desea enunciar, pero en la medida en que reciba realimentación positiva o negativa de sus interlocutores, en que éstos compitan por la posesión de la palabra y de cómo evolucionen las circunstancias sociolingüísticas que rodean a su producción, podrá verse obligado a reformular dinámicamente la porción de intervención pendiente. Tampoco podrá evitar que, en cualquier momento, algún otro participante pueda competir por la posesión de la palabra (comenzar a hablar solapadamente), mostrarse como candidato a poseerla (solicitar la palabra) o que la evolución de las circunstancias le haga receptor de la misma (por alusiones o eventos que le conviertan en centro de la interacción). Del mismo modo, las alternativas de contribución de un participante en la interacción no se restringen a intervenciones (contribuciones de pista de acción primaria), sino que en la interacción tiene también cabida la comunicación colateral [Apartado 2.2.3.2], que no requiere de la posesión de la palabra para ser producida. Todas estas posibles situaciones podrían alterar la formalización preliminar de la intervención del hablante, provocando una interrupción, obligándole a concluir de forma prematura, o a redefinir su discurso para mostrar mayor dominancia sobre la palabra para retenerla. Y dado que el

desenlace de estas posibles situaciones no está determinado, tampoco es posible saber de antemano cuál será el orden de intervención de los participantes en la interacción.

En definitiva, la interacción natural, en contra de la visión tradicional, es una acción colaborativa en la que cada acción es desarrollada conjuntamente por todos los participantes [Apartado 2.2]. Ésta es construida tanto por las intervenciones de los hablantes, como por las evidencias de cierre que los oyentes los aportan a través de contribuciones de realimentación. Como resultado, ni la duración de las intervenciones, ni su orden o contenido están definidos de antemano, sino que surgen dinámicamente según los participantes entiendan que deben o no contribuir a la interacción en un momento dado, incluso a pesar de producir solapamientos o interrupciones. Una toma de turnos restringida al *ciclo de interacción* [Apartado 2.1] constituye una importante limitación para un desarrollo natural de la interacción, que se hace más pronunciado a medida que las capacidades interactivas de los sistemas mejoran en aspectos como la pro actividad, la adaptabilidad a las circunstancias sociolingüísticas y la gestión del compromiso. Por todo ello, y de cara a alcanzar un comportamiento temporal más natural, los Sistemas de Interacción Natural deben estar sustentados en arquitecturas basadas en las *teorías de la acción combinada*, e implementadas sobre plataformas que permitan desarrollar una correcta independencia de procesos.

Capítulo 3 SISTEMAS DE INTERACCIÓN NATURAL

A lo largo de este trabajo se entenderá por Sistema de Interacción Natural cualquier sistema con aspiraciones a ofrecer una interfaz de comunicación entre las personas y la tecnología a través de códigos, modos y procedimientos humanos. Aunque, hoy por hoy, no puede hablarse de sistemas capaces de reproducir de forma completa las habilidades interactivas humanas, si existe gran diversidad de sistemas que, al menos de forma parcial, se acercan a este comportamiento. Los avances en este campo se producen a gran velocidad y en muy diversas direcciones, y los resultados obtenidos tienen cada vez una mayor repercusión, tanto en el ámbito académico como en el comercial.

Aunque el verdadero desarrollo de los sistemas con aspiraciones a reproducir la interacción humana pertenece a las últimas décadas, existen algunas aproximaciones previas que merecen ser tomadas en consideración. Quizás el caso más antiguo sea la máquina de Von Kempelen [74], un sistema capaz pronunciar palabras y oraciones cortas (de mayor o menor complejidad en función de la pericia del operador), lo que se conseguía al hacer circular aire a través de unos conductos y cavidades pensados para modular una veintena de sonidos distintos. Fue la aparición de la electricidad la que permitió realizar los primeros avances notables en este tipo de tecnologías. Homer Dudley presentó en 1939 [43] un sistema eléctrico denominado VODER que aplicaba *codificación predictiva lineal* (LPC) para representar la voz. No obstante, la verdadera revolución en la síntesis de habla llegó de la mano de Frank Cooper [39], quien haciendo incidir un rayo de luz sobre un disco giratorio perforado, conseguía modular una señal luminosa y transformarla en sonido al proyectarla sobre un espectrograma de células fotovoltaicas.

En los años 60, con la llegada de los ordenadores, se llevan a cabo las primeras investigaciones en reconocimiento de habla, y se consiguen resultados a nivel de fonemas. En paralelo comienzan a implementarse algunos sistemas de conversación por texto. Destaca el Sistema ELIZA [189], que tenía asociados mensajes de respuesta para cada posible mensaje del usuario. El objetivo de estos sistemas era la imitación de las conversaciones humanas, dejando de lado cuestiones como la interpretación del mensaje o la gestión del diálogo.

Durante los años 70 se desarrollan sistemas como SHRDLU [193] y, con ellos, comienza a ser posible la comprensión del lenguaje natural. A partir de los años 80, el desarrollo de los sistemas de interacción gana fuerza y se trabaja en hacer más humana la interacción. Aunque la tecnología de reconocimiento de habla aún no soporta reconocimiento de enunciaciones completas, se trabaja en la interpretación, gestión y generación de expresiones largas. De esta década pueden destacarse los sistemas HAM-ANS [179], EES [164] y MINDS [197].

En los 90 aparecen las primeras arquitecturas conceptuales de Sistemas de Interacción Natural. Así comienzan a ser estructuradas con mayor precisión cada una de las tareas involucradas en la *interacción natural*. Comienza a hablarse de aspectos tales como el *modelado del diálogo* y, al mismo tiempo, la interacción pasa de ser un fin en sí mismo para estos sistemas, a convertirse en un medio para la resolución de otras tareas (por ejemplo, la consulta de información). Se trata principalmente de sistemas en los que la conversación es dirigida por el sistema y se desarrolla en pocos turnos. Se aplica a dominios como el seguimiento de paquetes o el estado de los vuelos. También comienza a ser una realidad el reconocimiento de enunciaciones habladas. Los sistemas PLUS [13], SUNDIAL [3], ARTIMIS [150], TRAINS [1], VERBMOBIL [145] o TRINDI [171] pertenecen a este periodo.

Entre el 2000 y el 2005 la interacción se aplica fundamentalmente a dominios transaccionales (banca, control de inventario comercial, reserva de trenes, etc.) en los que los intercambios llegan hasta la decena de turnos y las enunciaciones ganan en naturalidad. Se desarrollan las primeras versiones de la recomendación Voice eXtensible Markup Language, VoiceXML, [177] del World Wide Web Consortium (W3C) para la especificación de diálogos interactivos entre personas y ordenadores.

Los sistemas implementados desde ese momento hasta la actualidad se orientan a la resolución de problemas más amplios que la consulta de información o la realización de transacciones. Hoy en día, su aplicación en ámbitos comerciales se encuentra en plena expansión, y es ya una realidad que se extiende a dominios de aplicación muy diversos (estando incluso generalizada en muchos de ellos). De entre todos, pueden destacarse los siguientes:

- Atención al cliente
- Asistencia técnica
- Reservas de hoteles, vuelos o alquiler de vehículos
- Sistemas interactivos en el coche
- Monitorización de pacientes y asistencia sanitaria
- Educación y tutores inteligentes
- Videojuegos

El presente capítulo pretende ofrecer una visión global acerca de este tipo de tecnologías. Serán descritas las clasificaciones más aceptadas de Sistemas de Interacción Natural, y se prestará una particular atención a aquellos ejemplos que, por incorporar independencia de procesos o mejoras en la toma de turno, sean considerados especialmente relevantes. Por último, se revisarán las metodologías existentes para la evaluación de Sistemas de Interacción Natural.

3.1 *CLASIFICACIÓN DE LOS SISTEMAS DE INTERACCIÓN NATURAL*

Aunque los Sistemas de Interacción Natural admiten diversas clasificaciones (de acuerdo a su aplicación, dispositivo sobre el que se aplican, estilo de interacción, etc.), según Cohen [38], Mc Tear [127] y López-Cózar [122], las principales son las que atienden a los siguientes criterios: El papel que desempeña el sistema como agente interactivo; las modalidades y códigos utilizados; la dirección de la iniciativa; y el modelado del diálogo. Estas clasificaciones serán revisadas en los próximos apartados.

3.1.1 Clasificación Según el Papel que Desempeña el Sistema como Agente Interactivo

En función del papel que desempeña el sistema en la interacción, pueden distinguirse distintas categorías de Sistemas de Interacción Natural. En primera instancia, cabe considerar si el sistema es un mero recurso para la obtención de información y la resolución de tareas, o si posee en la interacción entidad propia como participante, en la misma medida que los usuarios humanos con los que interactúa. Además, deberá considerarse si sus capacidades interactivas se limitan a la mera imitación del comportamiento humano (sin pretensiones de modelar el conocimiento real puesto en juego en la interacción) o si, yendo más allá, pretende modelar en mayor profundidad los sus mecanismos interactivos.

Partiendo de estos dos criterios, y basándose en la clasificación de Michael Mc Tear [127], se pueden distinguir los siguientes tipos de Sistemas de Interacción Natural:

- *Sistemas orientados a tarea*: Estos sistemas se utilizan como interfaz para extraer información de bases de datos o para llevar a cabo transacciones.

- *Sistemas conversacionales*: Son los sistemas que dan tanta importancia a la interacción como a la propia resolución de las tareas. A su vez, puede distinguirse entre:

    - *Sistemas de conversación simulada*: El objetivo es simplemente la imitación de las habilidades interactivas humanas.

    - *Sistemas discursivos*: Conciben la conversación como una actividad en la que participan agentes con sus propias creencias, deseos e intenciones.

3.1.1.1 Sistemas Orientados a Tarea

Los sistemas orientados a tarea desarrollan la interacción según la *metáfora de rellenado de formularios*. Son sistemas de iniciativa dirigida por el sistema. En ellos, este va solicitando al usuario los valores de cada uno de los campos requeridos para realizar la tarea y el usuario responde a cada pregunta por orden. Una vez obtenida toda la información necesaria, el sistema realiza la tarea devolviendo al usuario la información solicitada o confirmando la transacción. Resultan adecuados cuando la naturaleza de la conversación a desarrollar se adapta bien a este paradigma interactivo (por ejemplo, en el campo de la banca telefónica), pero producen interacciones extremadamente rígidas e inflexibles. Estos sistemas son muy comunes hoy en día en el ámbito comercial, y están ampliamente extendidos entre los sistemas telefónicos de atención al cliente, banca telefónica, o sistemas de reserva de billetes.

3.1.1.2 Sistemas de Conversación Simulada

Los sistemas conversacionales tienen como objetivo principal el desarrollo de interacciones lo más similares posible a las que desarrollan las personas entre ellas. La interacción con la tecnología tradicionalmente ha requerido que las personas realicen mayor esfuerzo en asimilar los códigos, procedimientos y modos que de los distintos dispositivos con los que se comunican y el reto en este caso es invertir el esfuerzo que realiza cada una de las partes, y hacer que la tecnología sea la que se adapta a la interacción que a las personas les resulta natural.

Dentro de este grupo se encuentran los sistemas de conversación simulada (como los *compañeros conversacionales*, los *chatbots*, *sistemas de ayuda de escritorio*, etc.). Son sistemas que tienen como propósito simular conversaciones de apariencia lo más humana posible, pero sin realizar una comprensión profunda de la interacción, ni modelar el diálogo. Sólo pretenden emular el comportamiento conversacional humano y, en general, son sistemas con iniciativa dirigida por el usuario que suelen limitarse a encajar sus enunciaciones en patrones a los que se les asigna una determinada respuesta. Su utilización está muy extendida como herramienta complementaria a páginas Web de ámbito diverso, (especialmente de comercio electrónico), como alternativa para la consulta de catálogos y la ayuda en línea. Algunos ejemplos tradicionales son ELIZA [189] o PARRY [124], aunque actualmente se pueden encontrar otros muchos ejemplos. Destacan los aplicados al ámbito comercial como el sistema Anne (http://193.108.42.79/ikea-es/cgi-bin/ikea-es.cgi), el agente de ventas virtual de la cadena de tiendas de muebles IKEA.

3.1.1.3 Sistemas Discursivos

Por su parte, los sistemas discursivos, son sistemas que aplican técnicas de inteligencia artificial para desarrollar capacidades conversacionales avanzadas. Se apoyan en conocimientos propios de la lingüística, sociología o psicología. Por ello consideran que la conversación es desarrollada por agentes interactivos (independientemente de que se trate del sistema o de usuarios) que desarrollan la interacción en base a sus propias creencias, sus deseos e intenciones. Bajo esta perspectiva, la conversación solo puede llevarse a cabo si el sistema es capaz de entender la interacción. Algunos ejemplos de sistemas discursivos son Communicator [185], TRIPS [2] o JASPIS [172].

Conseguir que la interacción entre las personas y la tecnología se desarrolle con naturalidad pasa por conseguir que los sistemas: (a) se comporten y expresen de tal forma que los potenciales usuarios puedan percibir, identificar, entender y reconocer las motivaciones que subyacen a sus expresiones, sin necesidad de realizar mayor esfuerzo que el que se requeriría si la enunciación fuese realizada por otra persona; (b) perciban, identifiquen, entiendan y reconozcan las motivaciones de la enunciaciones de los usuario, tal y como lo haría otra persona.

El sistema debe ser capaz de interpretar en qué forma afectan al contexto y al estado de la interacción las enunciaciones de los usuarios y, al mismo tiempo, de producir enunciaciones adecuadas a esa situación y de entender cómo estas afectan también al contexto y al estado. Por ello, este tipo de sistemas aplican teorías como las de la acción combinada, las teorías de planes

o las del estado de la información. Según esta forma de entender la interacción, es crucial el correcto modelado del conocimiento relativo al diálogo, usuario, sesión, situación, emociones e incluso el propio sistema.

Este tipo de sistemas consumen una elevada cantidad de recursos para desarrollar su interacción, debido a que contienen una representación más profunda del estado en el que se encuentra la interacción y el conocimiento sociolingüístico asociado. Es por ello que, a pesar de tratarse de la aproximación interactiva de mayor potencial, no siempre es la más adecuada. Esto es así especialmente cuando los tiempos y costes de desarrollo son limitados o los dispositivos sobre los que deben montarse son restringidos. No obstante, en la medida en la que las habilidades interactivas de los Sistemas de Interacción Natural crecen, se hace también más necesario partir de este enfoque para el desarrollo de estos sistemas. Concretamente, para el tratamiento de una *toma de turno avanzada* se hacen indispensables las soluciones basadas en el enfoque discursivo, puesto que son requeridas, tanto la comprensión exhaustiva de la interacción, como la de su estado por parte de todos los participantes implicados (incluido el propio sistema).

3.1.2 Clasificación Según las Modalidades Soportadas

Según las modalidades y códigos que soportan, los sistemas de diálogo se pueden clasificar en [122]:

- *Monolingües*: Sólo utilizan como entrada y salida habla o texto en un único idioma.
- *Multilingües*: Permiten la interacción hablada o escrita a través de diferentes lenguas.
- *Multimodales*: Los usuarios pueden comunicarse con el sistema a través de diversos canales en entrada, salida o ambos. Entre los canales de entrada se pueden destacar voz, movimiento de los labios, gestos, miradas, muecas, etc. Entre los de salida voz, imágenes, gráficos, sonidos, etc.

3.1.2.1 Sistemas Monolingües

Los sistemas monolingües solo pueden desarrollar la interacción por habla o texto, y en una única lengua. Este tipo de tecnologías comenzaron a aplicarse en los años setenta y pertenecen a este grupo los sistemas ELIZA [189] o SUNDIAL [3], entre otros. De los tres tipos, son los sistemas más simples. Se componen de:

- Una interfaz de entrada que adquiere e interpreta la señal de voz o texto emitida por el usuario y la convierte a un flujo de datos semánticamente igual, pero comprensible por el resto de los componentes del sistema (frecuentemente aplicando las teorías de los actos de habla).
- Un componente de interacción que contiene, o bien los modelos de diálogo, o bien el conjunto de reglas y patrones que con la información de contexto permiten elaborar las respuestas del sistema.
- Una interfaz de salida que, a partir de las respuestas del componente de interacción (expresadas como un flujo de datos distinto al lenguaje natural), genera la enunciación de respuesta y la expresa como voz o texto.

3.1.2.2 Sistemas Multilingües

Los sistemas multilingües son los sistemas capaces de desarrollar la interacción en distintos idiomas. Para ello requieren interfaces más complejas, con respecto a los sistemas monolingües, tanto a la entrada como a la salida. La interfaz de entrada debe ser capaz, en primer lugar, de reconocer la lengua en la que el usuario emitió su enunciación. En algunos casos la selección del idioma debe ser realizada de forma manual por el usuario. Una vez identificado el idioma, debe reconocerse e interpretarse la enunciación. Para ello se requieren reconocedores e intérpretes específicos para cada una de las lenguas soportadas por el sistema. Por su parte, la interfaz de salida, requiere generadores de lenguaje natural específicos para cada lengua y también distintos sintetizadores de voz para cada una de ellas o sintetizadores multilingües (como Festival [35]). El componente de interpretación, por su parte, no requiere cambios importantes con respecto a las aproximaciones monolingües (por trabajar con las estructuras semánticas independientes del idioma) y su única nueva función será la de contribuir a la selección del idioma en que se expresarán los mensajes de salida del sistema. Algunos ejemplos de sistemas multilingües son Jupiter [199] y Voyager [68].

Este tipo de sistemas también pueden ser aplicados a la traducción automática. Dado que el procesamiento de la interacción se realiza sobre una representación semántica independiente de la lengua, pueden convertirse las enunciaciones de entrada del usuario a una lengua distinta aprovechando las ventajas que el modelado de la interacción aporta a la robustez y corrección de la traducción. El sistema VERBMOBIL [145] es un ejemplo de este tipo de aplicaciones.

3.1.2.3 Sistemas Multimodales

Los sistemas multimodales dan un paso más hacia la *interacción natural*. La *interacción humana* tiene una estructura triple básica (es decir, formada no sólo por el *lenguaje*, sino también por el *paralenguaje* y la *quinesia*) [142; 143] y, además del habla, contiene gestos, expresiones faciales, miradas, etc. que complementan, modifican o sustituyen la información introducida a través del canal de habla (bien sea en la presentación primaria o en alguna presentación colateral). La interacción humana es, por tanto, multimodal.

La introducción de la multimodalidad afecta a todos los componentes que integran los sistemas de interacción. La interfaz de entrada requiere, junto a los reconocedores de habla y texto del resto de tipos, reconocedores e intérpretes de las distintas modalidades que soporta el sistema: gestos, escritura manual, posición y orientación del cuerpo, de los ojos, movimiento de labios, etc. La interfaz de salida también requiere modificaciones similares y las versiones multimodales incorporan *componentes de generación gráfica* que habitualmente consisten en personajes virtuales o avatares bi y tridimensionales que gesticulan y se comportan como personas durante la interacción. Se cuidan aspectos como el movimiento de los labios, la expresión facial, la mirada, la dirección del cuerpo, la expresión con las manos, etc.

Así mismo, la representación de enunciaciones que maneja el sistema internamente ya no puede estar apoyada en las teorías de actos de habla (por limitarse al habla), y se requieren aproximaciones más amplias, como las teorías de los actos comunicativos, para soportar también la representación de las contribuciones pertenecientes a otras modalidades distintas.

La posibilidad de una interacción no restringida a una única modalidad hace más transparente, potente, flexible y efectiva la comunicación. Entre las diversas ventajas que aporta la multimodalidad se pueden destacar las siguientes:

- Una comunicación más robusta en entornos cambiantes, ya que se puede sacar provecho de la redundancia que ofrece la combinación de diversos modos (por ejemplo habla y movimiento de labios) para corregir errores y mejorar la confianza. Algunas estrategias que pueden ser aplicadas son los *mapas interactivos* de Oviatt [136].
- Conmutar de unos modos a otros cuando la privacidad es importante (por ejemplo, sustituyendo habla por texto) o cuando la situación dificulta la comunicación por alguno de ellos (por ejemplo, se puede evitar el habla en entornos ruidosos, o adaptar los modos utilizados a las características del

dispositivo utilizado). En estos casos el sistema puede, entre otras cosas, sugerir a los usuarios utilizar alguna modalidad alternativa.

- Conmutar de unos modos a otros por preferencias del usuario o ante algún tipo de discapacidad. Esto permite aumentar el rango de usuarios potenciales y mejora la adaptabilidad en combinación con perfiles de usuario a través del modelo de usuario.
- Facilita la desambiguación y la cooperación entre modalidades [100] para complementar la información y la resolución de determinadas expresiones referenciales (por ejemplo, decir "*pon eso ahí*", señalando a algún lugar en concreto) o para aportar redundancia (por ejemplo, decir "*dirígete hacia la izquierda*", señalando a la izquierda). Este tipo de enunciaciones que combinan contribuciones de distintas modalidades producidas simultáneamente permite realizar intervenciones más cortas que utilizando una modalidad única, de forma que la multimodalidad permite agilizar la interacción en gran medida.
- Estimulación simultánea de varios sentidos del usuario, lo que mejora el compromiso y la atención de los usuarios y permite ofrecer una mejor ayuda a los usuarios en la realización de determinadas tareas [108].
- Permite alcanzar a un mayor rango de aplicaciones y permite integrar otros modos de interacción como son el ratón, los formularios, cuadros de texto, botones, etc.

No obstante, el tratamiento de la multimodalidad en la interacción conlleva un aumento importante de la complejidad de los sistemas y es también el origen de problemas hasta ahora inexistentes. El más importante de todos es la posible aparición de contradicciones cuando a partir de una misma enunciación del usuario, los reconocedores de distintos modos devuelven interpretaciones que entran en conflicto. Además, la incorporación de nuevas modalidades puede repercutir en un aumento de la tasa de error en el reconocimiento. Esto sucede cuando la disponibilidad de nuevas modalidades da pie a los usuarios a emitir determinadas contribuciones de formas alternativas, a través de canales con menor tasa de éxito (reconocimiento de escritura, lectura de labios, interpretación de gestos, etc.).

Por estas razones, sólo es recomendable la incorporación de nuevas modalidades a los sistemas de interacción cuando los nuevos modos habiliten nuevas capacidades que antes no existían y, al mismo tiempo, mejoren radicalmente su usabilidad. Tal es el caso de la incorporación de habilidades interactivas relacionadas con la toma de turno y el desarrollo temporal de la interacción. Fenómenos de tal tipo son, en gran medida, gestionados a través del paralenguaje y la quinesia (entonación, gestos, miradas), y producidos con la participación de

modalidades alternativas al habla. Existen algunos ejemplos de sistemas con avances en el tratamiento y detección de este tipo de señales, como son los sistema REA [24], que es capaz de tratar interrupciones y gestos relacionados con el comportamiento de toma de turno, o FADE [139], que modela el estado de la palabra y detecta marcadores multimodales de toma de turno.

Otros sistemas multimodales destacables son SmartKom [178] (con multi pizarra), MATCH [96], Witas [113], y Pedestrian Navigation System [187]. La multimodalidad también ha sido aplicada en ámbitos comerciales a diversos dominios, destacando el de los videojuegos [84; 128] o la educación [161].

3.1.3 Clasificación Según la Dirección de la Iniciativa

Para que el diálogo progrese deben aparecer iniciativas en la interacción por parte de alguno de los participantes que proyecten nuevos turnos en el diálogo. La aparición de iniciativas en la interacción está asociada al inicio de nuevos intercambios, concretamente a la aparición de las primeras partes. En función del origen de la aparición de estos nuevos intercambios, se distinguen los siguientes tipos de diálogo:

- *Diálogos dirigidos por el usuario*: Es el usuario el único participante que puede introducir nuevos intercambios en la interacción.
- *Diálogos dirigidos por el sistema*: El sistema determina el curso de la interacción.
- *Diálogos de iniciativa mixta*: Cualquiera de los dos participantes (usuario o sistema) pueden iniciar nuevos intercambios en la interacción.

3.1.3.1 Diálogo Dirigido por el Usuario

En los diálogos dirigidos por el usuario, el usuario es el único participante con derecho a iniciar nuevos intercambios en la interacción. El sistema se mantiene a la espera de la aparición de la primera parte de algún nuevo intercambio y, cuando esta se produce, desarrolla junto al usuario el resto de turnos del intercambio.

A efectos prácticos, en este tipo de diálogos el usuario se presenta como el participante que realiza las preguntas y el sistema como el participante que las interpreta y responde. En estos sistemas recae una importante carga de la interacción en el procesamiento e interpretación de las enunciaciones del usuario, puesto que el rango de las posibles acciones de entrada del usuario es muy amplio. Para abordar este problema, se suelen plantear interacciones en las que el usuario tiene cierto conocimiento del tipo de expresiones y frases que el sistema es capaz de procesar e interpretar.

Un ejemplo clásico de este tipo de sistemas es SUNDIAL [3], aunque este tipo de aproximaciones de diálogo suelen aplicarse a los sistemas de enrutamiento de llamadas ("*call routing*") y a los de recuperación de información en forma de pregunta-respuesta ("*question answering*") [Ejemplo 4].

Usuario: ¿En qué año nació Velázquez?
Sistema: El 6 de junio de 1599

Ejemplo 4: Iniciativa dirigida por el usuario

Bajo esta aproximación el sistema se limita a seguir la dirección de la iniciativa del usuario, y sólo lo hace en caso de que ésta se produzca.

3.1.3.2 Diálogo Dirigido por el Sistema

El sistema determina el curso de la interacción. Sólo él puede iniciar los nuevos intercambios y el sistema se limita a esperar a que se den tales inicios para desarrollar la parte del diálogo que le corresponde.

Suele tratarse de aproximaciones en las que el sistema solicita información al sistema, en forma de preguntas, y el usuario la facilita en consecuencia. Este tipo de sistemas simplifican el procesamiento e interpretación de las enunciaciones de usuario (el sistema sabe qué esperar del usuario en cada momento), pero da lugar a interacciones menos flexibles.

Los actuales sistemas telefónicos de atención al cliente, banca telefónica y gestión de reservas, por ejemplo, siguen este paradigma [Ejemplo 5, en inglés].

System: What is your destination?
User: London.
System: Was that London?
User: Yes.
System: What day do you want to travel?
User: Friday.
System: Was that Sunday?
User: No.
System: What day do you want to travel?

Ejemplo 5: Iniciativa dirigida por el sistema [126, pp.93]

En estos casos, el sistema toma las riendas de la interacción y delimita el espectro de posibles respuestas del usuario. No se contempla la posibilidad de que el usuario redirija la iniciativa en función de sus intereses y metas personales.

3.1.3.3 Diálogo dirigido por Iniciativa Mixta

Cualquiera de los dos participantes (usuario o sistema) pueden iniciar nuevos intercambios en la interacción. Este planteamiento suele estar vinculada a los sistemas de enfoque discursivo, en los que todos los participantes, tanto usuarios como sistema, son entendidos como agentes interactivos. Según este modelo, cualquier participante puede proponer nuevas metas individuales en la conversación (lanzar preguntas, iniciar temas y pedir aclaraciones, etc.), y puede hacerlo en cualquier momento a lo largo de sus turnos de participación. Este enfoque permite una interacción más compleja y hace posible la resolución de problemas de más envergadura [Ejemplo 6, en inglés].

User: *I'm looking for a job in the Calais area. Are there any servers?*
System: *No, there aren't any employment servers for Calais. However, there is an employment server for Pasde-Calais and an employment server for Lille. Are you interested in one of these?*

Ejemplo 6: Ejemplo de iniciativa mixta [126, pp.94]

La toma de turno producida en la interacción humana es no marcada, lo que quiere decir que ni el contenido, ni la duración de las contribuciones, ni el orden de contribución de los participantes están definidos de antemano. Esto conlleva contemplar la posibilidad de una redirección de la iniciativa por parte de cualquiera de los participantes en cualquier momento (tanto sistema como usuarios). Por ello, del mismo modo que se requieren enfoques discursivos y de orientación multimodal, el desarrollo de una toma de turnos avanzada parte de considerar una iniciativa mixta en la interacción, tanto en lo que respecta a las metas desarrolladas por los participantes en la interacción, como a los momentos en que deciden desarrollarlas.

Algunos ejemplos de Sistemas de Interacción Natural de iniciativa mixta son CMU Communicator [195], TRIPS y JASPIS.

3.1.4 Clasificación según el Modelado del Diálogo

El modelado del diálogo hace referencia a la representación de los intercambios que el sistema puede desarrollar en la interacción. También a la representación de las metas que

desarrollan estos intercambios, el estado en el que se encuentran y la forma en la que la interacción está conectada con la tecnología a la que los usuarios pretenden acceder. Partiendo de la clasificación de Cohen [36] y Michael McTear [127], los modelos de diálogo pueden ser:

- *Gramáticas de diálogo*: El flujo del diálogo está completamente determinado antes de comenzar y consiste en un conjunto de estados que definen tareas a realizar y transiciones entre estados que determinan las acciones que pueden participar usuarios y sistema: Es un tipo de aproximación con diálogo dirigido por el sistema.
- *Marcos*: Diálogos orientados a tarea (por ejemplo el rellenado de formularios). Permite diálogos más flexibles, aunque también dirigidos por el sistema.
- *Gestión intencional del diálogo*: El flujo del diálogo está completamente abierto, y se desarrolla como medio para realizar tareas. Tanto los usuarios como el sistema pueden tomar la iniciativa y entran en consideración sus metas y el contexto en cada instante.
- *Modelos de acción combinada*: Presuponen que la interacción es una acción colaborativa entre los participantes, que deben cooperar en una acción combinada y llegar a un compromiso para alcanzar sus metas individuales.

Todos estos casos pueden ser considerados modelos simbólicos, dónde la gestión del diálogo es modelada de abajo a arriba, partiéndose de un análisis de corpus real para determinar las habilidades que requiere desarrollar el sistema en la interacción. Los sistemas implementados de esta forma basan su gestión de diálogo en el procesamiento simbólico de reglas, referencias a planes, creencias, etc. Sin embargo, existe una alternativa a este tipo de gestión de diálogo, los modelos estadísticos, donde la interacción se modela a partir de un cálculo de probabilidades. Tanto los modelos de diálogo simbólicos como los estadísticos serán descritos en mayor profundidad a continuación.

3.1.4.1 Gestión Basada en Gramáticas de Diálogo

La aproximación a la gestión del diálogo basada en gramáticas de diálogo, también denominada aproximación de estados finitos, parte de la observación de que en los diálogos se presentan regularidades secuenciales derivadas del desarrollo de intercambios tipados: las preguntas van seguidas de respuestas, propuestas de aceptaciones, etc. Por ello se presupone que el diálogo tiene una estructura que puede ser representada en forma de gramáticas [162] e implementada fácilmente como autómatas de estados finitos.

Las gramáticas de diálogo definen el conjunto de acciones que pueden ser realizadas en cada punto del diálogo. Esta aproximación combina el conocimiento sobre las tareas y sobre el diálogo, de forma que el diálogo se representa por estados y transiciones entre estados. Un estado representa una etapa del diálogo en la que se pone en juego determinada información y en la que se realizan determinadas acciones. Las transiciones están basadas en las acciones de los usuarios o el sistema (mensajes de salida del sistema, respuestas de usuario, etc.). El conjunto de caminos posibles descritos en la gramática es el conjunto de diálogos posibles que pueden ser realizados.

Este tipo de aproximación resulta fácilmente implementable y es un enfoque adecuado para diálogos dirigidos por el sistema. También permite estructurar la resolución de problemas como secuencias de preguntas bien definidas a priori y facilita el reconocimiento e interpretación de las enunciaciones del usuario (ya que lo que puede decir el usuario en cada momento está acotado y es fácilmente predecible). Además, permite una representación gráfica del diálogo que sirve de gran ayuda para modelar la interacción.

La gestión de diálogo por gramáticas ha sido el enfoque aplicado a muchos de los Sistemas de Interacción Natural clásicos, como SUNDIAL [44], RAILTEL [112] o LINLIN [98], y es el enfoque elegido mayoritariamente en las aplicaciones comerciales y estándares existentes para el desarrollo de gestores de diálogo (VoiceXML [177]). También el que aplican las plataformas de desarrollo de sistemas de interacción, como son el CSLU Toolkit [125], SALT [184] o TRINDIKIT [170].

Entre las principales desventajas de las gramáticas de diálogo se encuentran que producen diálogos demasiado mecánicos e inflexibles. Esto da problemas con aquellos diálogos que se desvían del camino predefinido, haciendo difícil realizar correcciones, soportar habla solapada o introducir información no predicha en el tiempo de diseño. Aunque el diseño de diálogos con múltiples caminos posibles puede subsanar en parte estos problemas, deriva en un crecimiento exponencial del número de estados y transiciones que contienen los autómatas. Todo ello hace de esta aproximación poco apropiada para tareas complejas como, por ejemplo, la negociación (ya que no pueden desarrollarse diálogos cuyo curso no haya sido definido de antemano). Por ello, a pesar de sus ventajas, no es la aproximación más apropiada desde el punto de vista de la naturalidad de la interacción o, en cualquier caso, necesita funcionar conjuntamente con otros mecanismos para solucionar todos estos problemas.

3.1.4.2 Gestión Basada en Marcos

La gestión del diálogo por marcos es una aproximación declarativa. En esta aproximación la interacción consiste en el rellenado de unos formularios de campo denominados marcos. Cada marco está asociado al desarrollo de una tarea y define la información que se requiere para poder completarla. El desarrollo de un marco consiste en el rellenado de todos sus campos, y una vez que se completa puede realizarse su tarea asociada. El orden en el que se rellenan los campos de un marco no está predefinido y se decide en tiempo de ejecución en base a algoritmos simples. El sistema decide la siguiente cuestión que debe ser preguntada basándose en la información que ya ha recabado y la que le falta por recabar (de toda la que define el marco). El usuario, por su parte, no tiene por qué restringirse a rellenar un marco en el orden en que propone el sistema y podría aportar más información incluso de la que se le hubiera preguntado hasta el momento (y que intuye o presupone que será relevante o que el sistema preguntará en el futuro). Algunos ejemplos de que aplican este tipo de gestión del diálogo son los sistemas descritos por Wang [183], Bobrow [14] o el sistema de traducción automática VERBMOBIL [145].

Entre sus ventajas está el hecho de que es una buena aproximación para aquellos casos en los que el dialogo puede ser dirigido por las necesidades de información y también en los casos en los que las acciones pueden ser ejecutadas en ordenes distintos (por ejemplo, para realizar reservas de hoteles). Además, el marco permite estructurar por sí mismo el contexto del diálogo. En general, este tipo de aproximaciones permiten mayor flexibilidad en la interacción que las gramáticas de diálogo. Sin embargo, como principal desventaja se tiene que, con respecto a las gramáticas de diálogo, complica el proceso de procesamiento de lenguaje natural, la interpretación de las enunciaciones y la forma en la que progresa la interacción. Al imponer menos restricciones en lo que el usuario puede decir, requiere gramáticas más cercanas a las del lenguaje natural donde, además, deben ser contempladas las distintas permutaciones en que el usuario puede presentar la información requerida por cada uno de los marcos. Así mismo, el curso que puede seguir el diálogo no está predefinido y se requieren algoritmos de control más elaborados.

3.1.4.3 Gestión Basada en Modelos Intencionales

Los modelos intencionales entienden la interacción como un intercambio llevado a cabo entre agentes racionales para desarrollar tareas. Cuando uno de los participantes requiere la realización de determinadas tareas, se desarrollan entre los participantes los intercambios que permiten completarlas. De esta forma, cualquier participante puede introducir en la interacción

la necesidad de desarrollar nuevas tareas, por lo que se trata de una aproximación de iniciativa mixta. En el diálogo así entendido, la selección de acciones se realiza en base a reglas de comportamiento y metas. Este enfoque combina las intenciones de los hablantes y el espacio focal con las tareas de las metas subyacentes. Usa técnicas de inteligencia artificial (negociación, resolución de problemas, planificación, etc.) que idealmente deberían ser independientes del dominio de interacción.

Partiendo de los estudios de Grosz y Sidner [81; 80], el diálogo (el discurso) puede modelarse en diferentes niveles que son independientes, pero que están interconectados. Son los niveles intencionales, atencionales y lingüísticos. La gestión del diálogo debe mantener la coherencia del diálogo en un ámbito global en base a la estructura de las intenciones y las tareas y en ámbito local por la pila del foco. Este tipo de modelos contempla el desarrollo de las tareas como recetas cuyo desarrollo es introducido de forma unilateral por uno de los participantes. No se concibe la posibilidad de que el resto de participantes puedan rechazar su desarrollo. Esta aproximación ha sido implementada en sistemas como COLLAGEN [146], en el marco de trabajo de un interfaz de agentes colaborativos.

3.1.4.4 Gestión Basada en Modelos de Acción Combinada

Al igual que en la aproximación anterior, este enfoque parte del profundo análisis de la conversación entre personas para aplicar los mismos principios sobre la comunicación persona-ordenador. En este caso, los participantes también son entendidos como agentes racionales que se involucran en la interacción con el fin de alcanzar sus propias metas, pero en este caso, los modelos de acción combinada entienden la interacción como una acción colaborativa, en la que los planes se desarrollan con el acuerdo entre los participantes [97]. En este caso, se pone un mayor énfasis en la colaboración cooperativa que en el desarrollo de las propias tareas. Según este enfoque, el diálogo es construido como una secuencia de estados de interacción en el que cada nuevo movimiento supone el desarrollo de una o varias metas compartidas, o bien la inserción de nuevas metas individuales (que podrán ser aceptadas convirtiéndose en ese caso en metas compartidas). Bajo este enfoque los participantes ya no asumen sin más el desarrollo de los intercambios propuestos, aceptan su desarrollo sólo por voluntad propia y siempre cabe la posibilidad de que lo desestimen. TRIPS [2], JASPIS [172] o SOPAT [41] parten de esta aproximación.

Para desarrollar una *toma de turnos natural* a la interacción humana es imprescindible modelar la interacción en términos de acción combinada, puesto que la forma en la que ésta se produce a lo largo del tiempo surge del compromiso alcanzado entre las metas compartidas por

los participantes y de su evolución, así como de los equilibrios de estas metas compartidas con las metas propias que mantienen en cada momento los distintos participantes de la interacción. De esta forma, las arquitecturas que parten de las teorías de la acción combinada se presentan como la mejor alternativa para soportar habilidades avanzadas de toma de turno.

3.1.4.5 Modelos de diálogo estadísticos

Los modelos de diálogo estadísticos no se apoyan en el manejo de símbolos para desarrollar la gestión del diálogo, sino que aplican como estrategia interactiva el cálculo probabilístico. Se trata de sistemas capaces de aprender los mecanismos de interacción aplicando procesos de aprendizaje automático a partir del corpus previamente recogido. Este tipo de modelos de diálogo evitan, en gran medida, el uso de reglas explícitas impuestas a priori, pero conllevan una fuerte dependencia del corpus utilizado para el entrenamiento. Por ello son altamente sensibles al dominio. Las alternativas más comunes para su modelado son: MDP (Markov Decission Process); POMDP (Partially Observable Markov Decission Process); Redes Bayesianas; clasificadores basados en n-gramas e inferencias gramaticales; y Redes Neuronales.

Los gestores de diálogo basados en MDPs requieren, además de los módulos de interpretación, generación y modelo de diálogo, módulos de recompensa, optimización y decisión. El modelo de diálogo, en este tipo de aproximaciones, representa explícitamente los siguientes elementos:

- El espacio de estados contiene todos los posibles estados que puede alcanzar la interacción.
- Conjunto de acciones que el agente puede realizar potencialmente en cada estado.
- Funciones de recompensa asociadas a cada estado que capturan las consecuencias de la elección de estados realizada para producir recompensas positivas o negativas.
- Transiciones probabilísticas al estado siguiente determinadas por elección de acciones en el estado actual.

El objetivo es encontrar una política que refleje la mejor acción a tomar en cada estado utilizando aprendizaje por refuerzo. El aprendizaje por refuerzo permite optimizar las prestaciones del sistema por la exploración sistemática de todas las distintas acciones que el sistema puede tomar y evaluando cual de todas ellas es la más adecuada en cada momento (según una función de utilidad). El sistema aprende al ser recompensado cuando hace elecciones que le ayudan a obtener una mejor salida. La mejor política es la que tiene la mejor recompensa

estimada sobre todas las posibles secuencias de estados del diálogo. Un caso particular son los modelos de diálogo basados en POMDPs, en los que la creencia no se restringe a una hipótesis única sobre el estado de interacción, sino que mantienen una función de distribución sobre un conjunto más amplio de posibilidades. Destacan en esta línea los trabajos de Thomson y Young [168] y Williams [192].

Algunos trabajos [56; 79] profundizan en la aplicación de motores de inferencia basados en Redes Bayesianas para mejorar en la identificación de los denominados *objetivos de diálogo* (intenciones) a partir de la información semántica aportada por el usuario en sus enunciaciones y del contexto. Entre las ventajas de este tipo de motores de razonamiento, destaca la posibilidad de realizar análisis de congruencia entre las intenciones del usuario que el sistema estima a partir de sus enunciaciones y el contexto recogido durante la interacción, lo que mejora la reacción del sistema de acuerdo a la lógica del dominio de interacción. La principal ventaja es que, a partir de un entrenamiento máquina, es posible detectar de forma automática qué conceptos son necesarios, erróneos u opcionales en relación a los objetivos inferidos. De este modo, el diálogo podrá dirigirse hacia la producción de mensajes solicitando aquellos términos que sean precisos, aclarar los erróneos y obviando los opcionales.

Los clasificadores basados en n-gramas e inferencias gramaticales [65; 160] toman un conjunto de muestras de entrenamiento y generan un autómata de estados finitos probabilístico para cada uno de los posibles estados de la interacción. Las transiciones de estos autómatas vienen determinadas por secuencias de cadenas sobre el alfabeto original de entrada, las cuales son etiquetadas aplicando funciones que dependen del tipo de clasificación. En los clasificadores basados en n-gramas, la etiquetación se realiza a partir de la posición de las palabras recibidas, mientras que en los modelos MGGI (Morphic Generator Grammatical Inference) se toman en consideración otras características de carácter sintáctico. Cuando se recibe una nueva muestra, ésta es analizada sobre cada uno de los autómatas aprendidos, seleccionándose como respuesta la asociada al estado de aquel autómata que proporciona la mayor probabilidad de acierto. Esta gestión de diálogo, en definitiva, consiste en la aplicación de lenguajes regulares obtenidos a partir de un conjunto de muestras positivas a la identificación del estado siguiente que se alcanzará con mayor probabilidad con cada muestra recibida (y que determina la respuesta que deberá generar el sistema). Estos modelos son altamente eficientes, aunque con un coste computacional de aprendizaje elevado.

Por su parte, las Redes Neuronales [79] organizan en capas un conjunto de neuronas (elementos de procesamiento simple) que se encuentran conectadas entre sí. Cada neurona está conectada con otras neuronas mediante enlaces de comunicación con un peso asociado. Es en

estos pesos donde se encuentra el conocimiento que tiene la red acerca de un determinado problema. Para la aplicación de una red neuronal en la búsqueda de respuestas en la gestión del diálogo, la capa de entrada recibe la codificación del estado de interacción y del estado actual, y como salidas define el conjunto de posibles estados siguientes o respuestas del sistema, representando la probabilidad de pertenencia de cada una de las muestras a cada una de las salidas. La principal ventaja de la aplicación de redes neuronales a la gestión del diálogo es su capacidad generalizar y manejar información difusa, puesto que mejoran conforme adquieren más conocimiento sobre el problema.

Este tipo de aproximaciones estadísticas tienen como principal ventaja, con respecto a las aproximaciones simbólicas, el facilitar el proceso de adaptación del gestor de diálogo a nuevos dominios de interacción. La razón está en que dichos modelos permiten automatizar la implementación de corpus a partir de un conjunto de muestras de entrenamiento. Esta automatización hace posible reducir el tiempo y los costes de desarrollo de dominios limitados, aunque en la medida en que el tamaño del corpus aumenta, el número de muestras de entrenamiento necesarias para mantener una tasa de éxito similar a la que ofrecen los sistemas simbólicos crece exponencialmente. Del mismo modo, las aproximaciones estadísticas no permiten la ampliación o modificación del corpus una vez entrenado, y ante la necesidad de adaptar el corpus a nuevos dominios de interacción se requiere un reentrenamiento completo de todo el sistema. Por su parte, los sistemas simbólicos sí permiten un desarrollo ágil e incremental del corpus, haciendo posible desarrollar independiente el conocimiento interactivo a aplicar sobre distintos aspectos del dominio de interacción, y su posterior recombinación en unidades mayores.

3.1.5 Conclusiones

Los Sistemas de Interacción Natural admiten clasificaciones en base a diversos criterios [Tabla 5]: i) según el papel que desempeña el sistema como agente interactivo, pudiendo ser orientados a tarea, de conversación simulada o discursivos [Apartado 3.1.1]; ii) según las modalidades y códigos utilizados, encontrándose sistemas monolingües, multilingües o multimodales [Apartado 3.1.2]; iii) por la dirección de la iniciativa, distinguiéndose entre sistemas dirigidos por el usuario, por el sistema o de iniciativa mixta [Apartado 3.1.3]; y iv) según la forma en la que modelan el diálogo, existiendo aproximaciones por juegos de diálogo, marcos, modelos intencionales, modelos de acción combinada [Apartado 3.1.4] y modelos estadísticos [Apartado 3.1.4.5Modelos de diálogo estadísticos].

Según estas clasificaciones, un desarrollo temporal de la interacción más cercano al desarrollado por las personas estaría enfocado a sistemas *discursivos*, *multimodales*, de *iniciativa mixta* y *acción combinada*, respectivamente. Se requieren sistemas discursivos ya que la forma y el momento en que son producidos los turnos en la interacción derivan de su profunda comprensión, así como de las circunstancias que la rodean. Multimodales porque la toma de turno de la interacción natural se gestiona en gran medida a través del *paralenguaje*. De iniciativa mixta dado que en la interacción natural la iniciativa de añadir, eliminar o hacer progresar las metas puede surgir en cualquier momento y provenir de cualquiera de los participantes. Finalmente, se requieren soluciones basadas en las teorías de la acción combinada porque, en última instancia, es el equilibrio entre las metas de los distintos participantes y el compromiso alcanzado sobre ellas lo que determina con qué urgencia deben los participantes tomar parte en la interacción, qué metas deben desarrollar al hacerlo y cómo son desarrollados, rectificados, reformulados o interrumpidos los turnos.

A lo largo del presente apartado se han revisado las clasificaciones de Sistemas de Interacción Natural más comúnmente aceptadas. Del mismo modo, se han identificado los tipos de Sistema de Interacción Natural que se perfilan como más adecuados para soportar una *gestión avanzada de turnos* en la *interacción natural*. Sin embargo, todas estas clasificaciones parten de aspectos relativos a la *organización local* y *global* de la interacción [Apartado 2.3], dejando de lado las cuestiones referentes a su *organización temporal* [Apartado 2.4]. Dado que, una vez cumplidos los requisitos descritos en lo referente a la organización local y global, es la organización temporal la que determina la forma en la que se producen y reparten los turnos en la interacción, será preciso complementar este estudio con el de las distintas estrategias desde las que se abordan los procesos de la interacción natural. Desde este punto de vista, podrá distinguirse entre sistemas que basan su desarrollo temporal en el *ciclo de interacción* (un único proceso constituido por una secuencia de fases) [Apartado 3.2] y sistemas que conciben la interacción como un conjunto de procesos autónomos e independientes que cooperan entre ellos para resolver de forma conjunta los problemas que entraña [Apartados 3.3; 3.4]. Si bien los primeros mostraran sin excepción una alternancia de turnos rígida y mecánica (a modo de partido de tenis), son los segundos los que permiten un reparto de turnos más complejo y en los que tienen cabida fenómenos como el solapamiento y la interrupción. De ellos se prestará especial atención a aquellos casos en los que se aborde, de algún modo, la gestión de turnos.

3.2 *ARQUITECTURAS DE CICLO DE INTERACCIÓN*

La arquitectura de Sistema de Interacción más madura es la que aborda la *interacción natural* como un proceso único en el que se suceden de forma cíclica una secuencia de fases (adquisición, interpretación, operación, generación y síntesis), tal y como describe el *ciclo de interacción* [Apartado 2.1]. Según este tipo de organización temporal, los turnos se alternan entre los distintos participantes a lo largo de toda la interacción y no es posible alterar el orden predefinido de los turnos, ni desarrollar fenómenos como el solapamiento o la interrupción. Según estas reglas, el reparto de turnos queda perfectamente definido y no se precisa una gestión adicional de turnos.

Tabla 5: Resumen de tipos de Sistemas de Interacción Natural

Criterio	Categoría	Descripción	Ejemplos
Papel del Sistema como Agente Interactivo	Orientados a Tarea	Desarrollan la interacción con el objetivo de recopilar los parámetros necesarios para ejecutar tareas y para devolver sus resultados	Sistemas telefónicos de diálogo para la atención al cliente
	Sistemas de Conversación Simulada	Imitación de las actividades interactivas humanas	ELIZA [189]
	Sistemas Discursivos	Los participantes, sistema y usuario, abordan la conversación como agentes con creencias, deseos e intenciones propias	TRIPS [2] JASPIS [172] SOPAT [41]
Modalidades y Códigos	Monolingües	Las contribuciones se producen en una única modalidad e idioma	SUNDIAL [3]
	Multilingües	Las contribuciones pueden producirse en distintos idiomas	VERBMOBIL [145]
	Multimodales	Las contribuciones son expresadas de forma coordinada a través de varias modalidades	SmartKom [178] MATCH [96]
Dirección de la Iniciativa	Dirigido por Usuario	El usuario es el participante que introduce las iniciativas en la interacción	SUNDIAL [3]
	Dirigido por Sistema	El sistema es el participante que introduce las iniciativas en la interacción	VoiceXML [177]
	Iniciativa Mixta	Ambos participantes, usuario o sistema, pueden introducir iniciativas en la interacción	TRIPS [2] JASPIS [172] SOPAT [41]
Modelado del Diálogo	Juegos de Diálogo	La interacción es descrita por conjuntos de estados y sus transiciones	SUNDIAL [3]
	Marcos	La interacción es descrita por formularios a ser rellenados	VERBMOBIL [145]
	Modelos Intencionales	La interacción se desarrolla con el objetivo de satisfacer las intenciones propuestas por los participantes	COLLAGEN TM [146]
	Modelos de Acción Combinada	La interacción consiste en el desarrollo de acciones combinadas que permiten a los participantes satisfacer sus propias metas individuales	TRIPS [2] JASPIS [172] SOPAT [41]
	Modelos Estadísticos	El modelo de diálogo se calcula a partir de un cálculo de probabilidades	Thomson y Young [168] Fernández Martínez [56]

Ejemplos clásicos de Sistemas de Interacción Natural que parten de este planteamiento son los sistemas ELIZA [189] y SUNDIAL [3], pero es especialmente relevante la plataforma VoiceXML [Apartado 3.2.1], una arquitectura de ámbito general creada para desarrollar de forma sencilla Sistemas de Interacción Natural a partir de documentos XML. VoiceXML cubre desde los aspectos de interfaz hasta los aspectos propios de la gestión del diálogo, y es una plataforma de gran aceptación tanto en el ámbito académico como comercial. Este ejemplo será analizado en mayor detalle a continuación.

3.2.1 VoiceXML

Hasta los años 90 el desarrollo de los Sistemas de Interacción Natural estuvo dominado por soluciones propietarias surgidas del ámbito académico como fruto de estudios relacionados con el reconocimiento del habla y la síntesis de expresiones en lenguaje natural. En la medida en que la tecnología ha ido ganando madurez, el conjunto de componentes y problemas que entran en juego con la *interacción natural* ha podido ir siendo estructurado y esto, unido al creciente interés comercial que ha tomado la Interacción Hombre-Máquina en los últimos años, estas soluciones propietarias han convergido en la familia de lenguajes Voice Extensible Markup Language (VoiceXML) [177].

En sí mismo, VoiceXML constituye un conjunto de recomendaciones del World Wide Web Consortium, amparado por empresas como AT&T, IBM, Lucent y Motorola y ampliamente extendido hoy en día en ámbitos comerciales. La plataforma VoiceXML dota a los desarrolladores de herramientas para la representación y tratamiento del conocimiento relacionado con los principales componentes de los sistemas de interacción. Entre sus objetivos está:

- Permitir una clara separación entre la interacción con los usuarios y los servicios lógicos que provee.
- Reforzar la interoperabilidad y la portabilidad (estableciendo un conjunto de capacidades comunes de plataforma y estandarizando los lenguajes para la representación de las gramáticas en las que basa el procesamiento de lenguaje natural (SRGS+SISR), la descripción de los mensajes de usuario (SSML), y la gestión del diálogo desde un enfoque de gramáticas y marcos con iniciativa tanto del usuario como mixta (VoiceXML), etc.
- Proteger a los desarrolladores de los sistemas de diálogo de los detalles a bajo nivel de la plataforma y la implementación de la arquitectura simplificando el desarrollo de este tipo de sistemas y su adaptación a nuevos dominios.

VoiceXML ha permitido realizar importantes avances en la usabilidad que ofrecen estos sistemas y cuenta también con gran cantidad de plataformas de desarrollo, entre las que se encuentran BeVocal Café (http://cafe.bevocal.com/), Loquendo Café (http://www.loquendo.com/es/services/loquendo_cafe.htm), Tellme Studio (https://studio.tellme.com/) o Voxpilot Voxbuilder (http://ode.voxpilot.com/index.php). Con todo esto, convierte a los Sistemas de Interacción Natural en soluciones atractivas para los usuarios y permite reducir tiempo y los costes de desarrollo. Es una alternativa ampliamente extendida entre la banca, el inventariado comercial, la reserva de billetes de trenes o aviones, o la atención al cliente y el soporte técnico.

A pesar de sus numerosas virtudes, esta arquitectura da poco margen a la resolución de los problemas relacionados con la *gestión avanzada de turnos*. Sus principales limitaciones en esta línea son:

- *Presentación según ciclo de la interacción*: La interacción con los sistemas que siguen la arquitectura VoiceXML consideran la interacción como el paso de un testigo, donde en cada momento solo un interlocutor participa, el hablante, quién está obligado a intervenir y decide unilateralmente durante cuánto tiempo lo hará. Por esa razón se imposibilita la participación simultánea de varios interlocutores, anulando toda posibilidad de ofrecer realimentación simultánea, co construcción de intervenciones, y eliminando toda opción de decisión sobre la toma de turno. Del mismo modo, en este planteamiento no tienen cabida los casos de reformulación o auto interrupción ante cambios en las circunstancias sociolingüísticas o por movimientos internos del sistema. Tampoco cabe reinterpretación de las enunciaciones del usuario.
- *Gestión de diálogo basada en gramáticas y marcos*: El modelado del diálogo según estos paradigmas no parte de la comprensión del uso que las personas hacen del lenguaje, sino de la mera representación de las estructuras que aparecen en él. Por ello, no tiene cabida bajo este planteamiento la correcta representación del equilibrio que las metas combinadas guardan entre ellas, ni del que guardan las metas propias de los interlocutores con las combinadas, relaciones de las que depende el orden con el que van surgiendo las participaciones (de forma solapada o no), su duración, y los contenidos desarrollados.
- *Imposibilidad de independizar los distintos procesos de la interacción*: La arquitectura VoiceXML considera la interacción como un proceso único, desencadenado ante la intervención del usuario. Todos los fenómenos

mencionados anteriormente (solapamiento, realimentación, reformulación, auto interrupción, reinterpretación y gestión de metas) requieren una separación de los distintos procesos que intervienen en la interacción natural.

Por ello, a pesar de que esta tecnología está ampliamente aceptada y que permite ahorrar tiempo y costes en el desarrollo de los Sistemas de Interacción Natural, para la resolución del problema que aborda este trabajo son necesarias otro tipo de arquitecturas en las que la interacción no se restrinja a una sucesión ordenada de turnos. Las más adecuadas serán, por tanto, aquellas que permitan independizar los diversos procesos que tienen lugar en la interacción natural.

3.3 ARQUITECTURAS MULTIPROCESO

La *interacción natural* no es un proceso secuencial en el que cada uno de los problemas involucrados pueda ser abordado en un punto concreto de la ejecución. En realidad, se trata más bien de un conjunto de procesos independientes desarrollados en paralelo, donde cada uno de los cuales se encarga de gestionar una parte concreta de los problemas y el conocimiento puesto en juego en su desarrollo [36].

Aunque las aproximaciones basadas en el *ciclo de la interacción* [Apartado 2.1] resultan adecuadas en numerosos escenarios, en la medida en la que entran en juego en la interacción la pro actividad del sistema y se contempla la representación del componente circunstancial y su evolución a lo largo del tiempo, se hace necesaria la aplicación de una *organización temporal* [Apartado 2.4] más compleja que el paso de testigo que define el ciclo de interacción. Por ello, los sistemas que pretenden abordar estas nuevas cuestiones de la interacción natural deben ser implementados sobre arquitecturas que permitan formalizar como procesos independientes los distintos componentes que las integran, y a la vez soporten su adecuada coordinación.

Las arquitecturas multiproceso se presentan como la mejor alternativa a la hora de soportar el tratamiento de la interacción natural. De entre ellas las Plataformas Multiagente, especialmente las de pizarra compartida, son las que ofrecen una mayor flexibilidad a la hora de desarrollar Sistemas de Interacción Natural con capacidad para soportar interacciones con una organización temporal avanzada.

3.3.1 Arquitecturas Multiagente

Por Plataformas Multiagente se entienden aquellas plataformas que disponen un entorno carente de inteligencia centralizada y que actúan por la cooperación de componentes autónomos e independientes denominados agentes. Aunque no existe un consenso para su definición, Wooldridge [194] propone el conjunto de propiedades necesarias en todo agente:

- *Autonomía*: los agentes actúan sin la intervención directa de humanos u otros agentes y tienen algún tipo de control sobre sus acciones y estado interno.
- *Habilidad social*: los agentes interactúan con otros agentes (e incluso con humanos) por medio de algún tipo de lenguaje de comunicación de agentes.
- *Reactividad*: un agente percibe su entorno y responde de forma apropiada en un tiempo razonable a los cambios que ocurren en él.
- *Pro actividad*: los agentes no actúan simplemente en respuesta a su entorno, sino que también deben exhibir un comportamiento dirigido por objetivos tomando la iniciativa.

A este conjunto, otros autores como Franklin y Graesser [62], añaden otros posibles rasgos:

- *Adaptatividad*: La capacidad de un agente para cambiar su comportamiento conforme al aprendizaje que el mismo realiza.
- *Benevolencia*: Un agente debe estar dispuesto a ayudar a otros agentes siempre que esto no le impida alcanzar sus propios objetivos.
- *Continuidad temporal*: Un agente se puede considerar como un proceso sin fin, que realiza sus funciones de forma continua.
- *Movilidad*: Si el agente puede trasladarse a través de una red telemática, para actuar en una maquina diferente a la de su propietario, por ejemplo, para buscar información en nombre de este.
- *Racionalidad*: Capacidad del agente para razonar a partir de los datos que recibe para obtener la mejor solución posible.
- *Veracidad*: Un agente no debe comunicar información errónea a propósito.

Entre las plataformas Multiagente más destacables se encuentran Java Agent Development framework (JADE), su extensión Jadex (JADE eXtension) y Open Agent Architecture (OAA).

Java Agent Development framework (JADE) [6] es una plataforma de *software libre* (bajo licencia LGPL) que simplifica la implementación de sistemas Multiagente y aplica los estándares de comunicación definidos por FIPA [61]. La plataforma se puede distribuir en diferentes máquinas independientemente del sistema operativo. JADE permite la implementación tanto de agentes deliberativos como reactivos (es una plataforma híbrida) e incluye dos agentes especiales llamados Agent Management System (AMS) y Directory Facilitator (DF) que facilitan la gestión de los agentes. El agente AMS facilita un servicio de información de todos los agentes registrados en la plataforma y el DF ofrece un servicio de páginas amarillas informando de los diferentes servicios registrados en la misma. Jadex (Figura 3) [16] extiende JADE con un potente motor de razonamiento y añadiendo a los agentes la representación de creencias (información que tiene el agente del entorno y de sí mismo), metas (motivaciones del agente y estados a los que quiere llegar) y planes (medios por los cuales se pretenden alcanzar la metas).

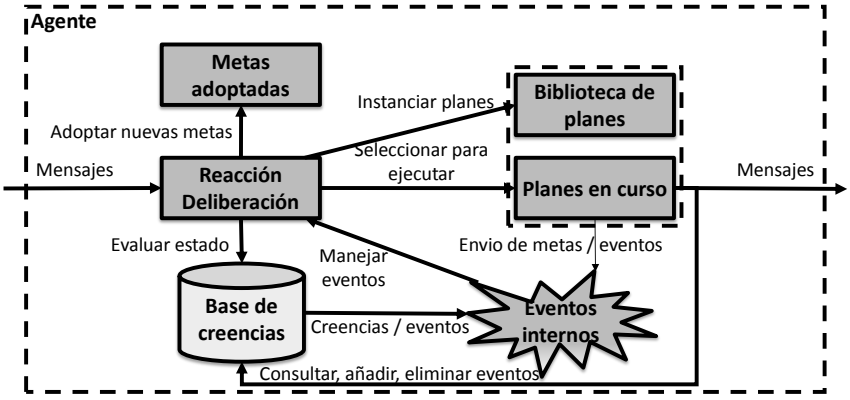


Figura 3: Arquitectura multiagente Extended Java Agent Development Framework (Jadex) [16]

Por su parte Open Agent Architecture (OAA) [29], tal y como muestra la Figura 4, utiliza como protocolo de comunicación Interagent Communication Language (ICL) y está estructurada en torno a un componente, denominado facilitador, que mantiene una lista de agentes proveedores de servicio y un conjunto de estrategias generales para alcanzar los objetivos (representados en los denominados meta-agentes). Los agentes solicitantes de servicio indican al facilitador las metas que pretenden alcanzar y este determinará cuál de los agentes proveedores es el más adecuado para resolverlo.

OAA ofrece una plataforma propia de desarrollo que permite reemplazar o añadir agentes en tiempo de ejecución y resulta muy adecuada para dispositivos poco potentes. Por estas razones OAA ha sido aplicada a distintos sistemas de diálogo, de los que MATCH [96] y Witas [113] son algunos ejemplos.

En la práctica, este tipo de arquitecturas son adecuadas cuando los servicios constan de un cliente y un único servidor y cuando se tratan de problemas con solución única. En muchos casos, en los procesos involucrados en la *interacción natural*, el mejor resultado se obtiene de la comparación de soluciones alternativas obtenidas competitivamente por diversas estrategias a partir de la misma solicitud. Del mismo modo, las relaciones entre los procesos no siempre quedan definidas en el momento de implementación, sino que surgen espontáneamente durante el desarrollo de la interacción, por lo que son necesarias otro tipo de soluciones.

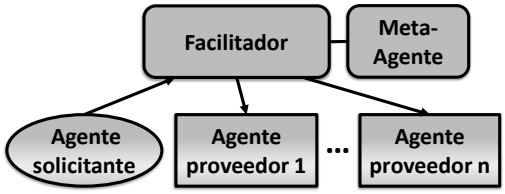


Figura 4: Arquitectura Multiagente Open Agent Arquitecture (OAA) [29]

3.3.2 Arquitecturas de Pizarra Compartida

Las plataformas de pizarra son aquellas en las que distintos procesos heterogéneos colaboran para resolver problemas compartiendo información a través de una pizarra: un repositorio de conocimiento común en el que se publican tanto las solicitudes como las respuestas.

Las aproximaciones de pizarra compartida, como FLiPSiDE [158] o LINDA [67], permiten la comunicación entre múltiples procesos por la lectura y escritura de información en un almacén de datos común a todos ellos. Los solicitantes publican sus solicitudes en la pizarra y esperan a la publicación de los resultados a dichas solicitudes. Por su parte, los servidores esperan a la aparición de solicitudes destinadas a ellos y, en ese caso, las atienden publicando sus respuestas en la pizarra.

La pizarra permite resolver problemas en equipo y obtener varios resultados para una misma solicitud, por lo que resulta ser una aproximación más flexible que la tradicional comunicación entre agentes. Además, este tipo de soluciones permite una composición flexible y dinámica de la plataforma, puesto que no obliga a que la interacción entre los componentes quede definida desde la propia implementación, sino que permite que pueda ser determinada en tiempo real. Algunos ejemplos de Sistemas de Interacción Natural implementados sobre arquitecturas de pizarra compartida son SmartKom [178], Pedestrian Navigation System [187] y SOPAT [41].

3.4 SISTEMAS CON GESTIÓN DE TOMA DE TURNO

La mayor parte de los Sistemas de Interacción Natural desarrollados hasta el momento presuponen que la palabra se pasa a modo de testigo de uno a otro participante a lo largo de la interacción, y de que lo hace en un orden predefinido. Se considera que el poseedor de la palabra la utilizará (y que lo hará durante el tiempo que considere oportuno) y que durante su turno ningún otro participante contribuirá simultáneamente. Este proceso de toma de turno es el que se ha venido a denominando *ciclo de interacción* [Apartado 2.1].

La toma de turno que se desarrolla en la *interacción humana* es, en realidad, un proceso mucho más flexible. Algunos sistemas que abordan una toma de turno más cercana a la humana son los Sistemas con Control de la Interrupción y los Sistemas con Gestión de la Realimentación. Los Sistemas con Control de la Interrupción (Barge-in Systems) [110] destacan por ser capaces de interrumpir la contribución del sistema cuando el usuario comienza a realizar una nueva contribución simultáneamente. Por su parte, los Sistemas con Gestión de la Realimentación [75;129;140] comprenden la importancia de ofrecer realimentación simultánea al usuario para participar en la coconstrucción de su intervención. En los Sistemas con Gestión de Realimentación actuales, las contribuciones de realimentación se desencadenan por cambios en la prosodia, movimientos de ojos o con los silencios. Ninguno de estos sistemas considera el estado en el que se encuentran las metas de la interacción, ni el compromiso alcanzado en torno a su desarrollo, ni el estado de la toma de turno, ni la influencia de las circunstancias sociolingüísticas. Por tanto, estos sistemas no pueden ser aplicados a la estimación de los momentos en los que el sistema puede participar en la interacción, bien sea según una contribución primaria o secundaria, y tampoco pueden considerar el efecto que la realimentación simultánea del usuario (entre otros tipos de contribución) debería tener en su propia contribución.

En general, un sistema con una toma de turno natural debería contemplar los siguientes aspectos:

- Independencia y coordinación de procesos.
- Interpretación y generación incremental, con mecanismos para gestionar la continuidad de las contribuciones.
- Representación del estado de las metas discursivas individuales del sistema y de las combinadas por los participantes.
- Tratamiento (interpretación y generación) de marcadores de toma de turno.

- Estimación del estado de los turnos, de la posesión de la palabra (a qué participante le toca hablar) y quiénes son los candidatos a tomarla.
- Mecanismos de decisión de toma de turno.

Aunque sin existir por el momento soluciones que aborden de forma completa un desarrollo natural de la toma de turno, si existen algunos trabajos que profundizan en la resolución de algunos de estos problemas. De ellos destacan los sistemas Ymir, FADE, VM-GEN y DECOP, que serán descritos con mayor profundidad a continuación.

3.4.1.1 Ymir

Thórisson [169] introduce una arquitectura para sistemas de diálogo multimodales reactivos que incorpora un modelo predictivo para estimar los momentos en los que le corresponde al sistema tomar el turno [Figura 5]. Para ello, propone una arquitectura multiproceso en la que simultáneamente es posible realizar la adquisición de las enunciaciones del usuario y la síntesis de las del sistema.

La clave de esta aproximación consiste en un procesamiento estructurado en tres capas diferentes, cada una de ellas con diferente prioridad de procesamiento. Estas capas son: capa de contenido (de baja prioridad), capa reactiva (de alta prioridad) y capa de control (de prioridad media). Además cuenta con un planificador de acciones, que determina qué procesos se ejecutan en cada momento.

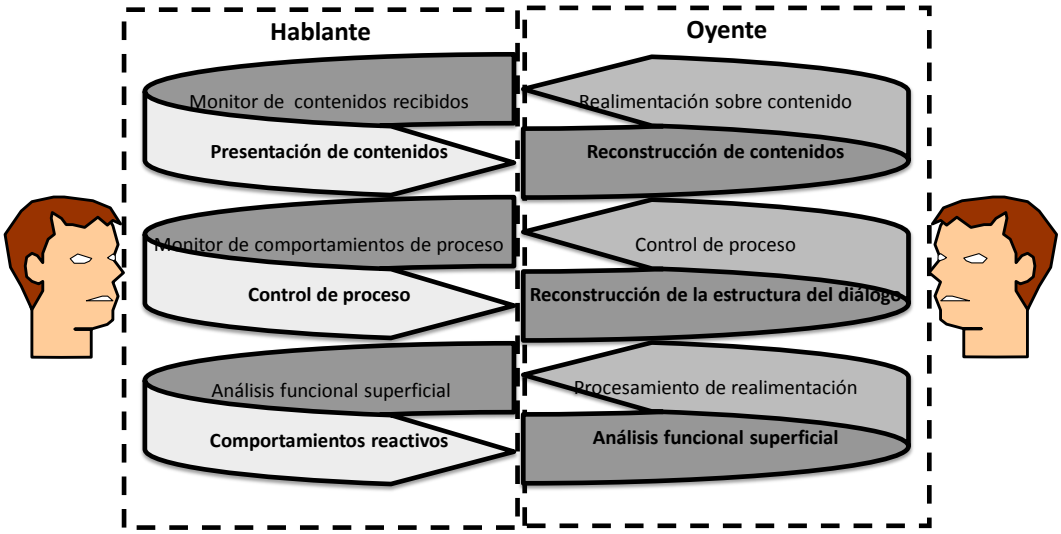


Figura 5: Arquitectura del sistema Ymir [169].

La capa de contenido es responsable de interpretar la entrada del usuario y de generar las respuestas apropiadas del sistema. En cuanto a la interpretación, la capa cuenta con unos

componentes denominados módulos de percepción, responsables de llevar a cabo la interpretación de las enunciaciones del usuario. Estos perceptores toman como entrada los datos de los sensores de adquisición (o de las salidas preprocesadas por otros módulos de percepción) y detectan las acciones de diálogo que contienen. Los resultados que producen son depositados en pizarras compartidas. En cuanto a la generación, este componente recibe de otras capas solicitudes de acción y, a partir de ella, determina las enunciaciones multimodales que producirá el sistema.

La capa reactiva se encarga de analizar los aspectos temporales de las acciones del usuario y, a través de los módulos de decisión reactivos, dispara acciones del sistema. La capa cuenta con módulos de decisión reactivos especializados en monitorizar el estado del diálogo, con módulos de decisión reactivos especializados en monitorizar el estado de los turnos y con módulos de decisión reactivos que toman decisiones sobre los comportamientos visibles de los participantes. Todos ellos toman como entrada los datos que los módulos de percepción depositan en pizarras compartidas.

Finalmente están la capa de control y el planificador de acciones. La capa de control es la responsable de activar y desactivar la ejecución de los distintos tipos de proceso, de representar el contexto de la interacción, y de gestionar el comportamiento interactivo del sistema. Por su parte, el planificador de acciones es el encargado de ejecutar las acciones generadas por el sistema.

Los logros de la propuesta relativos a una *toma de turno avanzada* son:

- Soportar la independencia de procesos de forma que pueden concurrir simultáneamente en el tiempo la interpretación de la enunciación del usuario y la generación de la del sistema.
- Una orientación multimodal que permite la detección de los marcadores de otros participantes relacionados con la toma de turno a partir de reglas (sólo de aquellos expresados de forma verbal o por gestos).
- La capacidad de estimar los momentos en los que le corresponde al sistema tomar turno, aplicando reglas de decisión.

Sin embargo, y a pesar de que permite interpretación y generación simultánea, no permite realizar estos procesos de forma incremental, por lo que entre ellos no existe una realimentación inmediata (sólo efectiva en futuras enunciaciones). Tampoco cabe posibilidad de realizar rectificaciones, reformulaciones o interrupciones. Además, la actitud del sistema de cara a la toma de turno es meramente pasiva, y se limita a estimar los momentos en los que el

usuario espera de él que participe. Tampoco es asumible por el sistema la iniciativa de intervenir por voluntad propia o de modificar en forma alguna la distribución de roles propuesta por el usuario. Además, la alternancia de turnos es dirigida por gestos explícitos (silencios, cesiones de turno, etc.), no considerándose aspectos como las alusiones. Finalmente, el sistema de toma de turno está concebido para una interacción exclusivamente bipartita. En definitiva, no se trata de una propuesta que aborde una toma de turno avanzada de forma global.

3.4.1.2 FADE

Norbert Pfleger [139] propone un sistema de fusión multimodal contextual, FADE (incluido en proyectos como Virtual Human [69], SmartWeb [133] o OMDIP [121]), que incorpora la capacidad de de interpretar y generar reactivamente señales relacionadas con la toma de turno, y también la de disparar reactivamente la toma de turno del sistema (tanto para intervenir como para ofrecer realimentación). El sistema parte de una orientación multiparticipante, y aplica una detallada clasificación de las acciones verbales y no verbales que afectan a la identificación de la toma de turno por parte del sistema y a la producción de la realimentación.

Su arquitectura [Figura 6] está compuesta por módulos de percepción (similares a los de la propuesta Ymir), módulos de contexto (que incluyen todo el conocimiento sociolingüístico asociado a la interacción), gestor de diálogo y generador multimodal.

El contexto se estructura entre contexto conversacional inmediato y contexto discursivo. En el primero se representa el contexto actual físico, perceptual y conversacional y en el segundo la contribución previa de los participantes individuales. Este último, además, representa en una memoria a largo plazo el conocimiento relacionado con el dominio de interacción (a través de una red semántica en la que a lo largo de la interacción van activándose los objetos que van siendo mencionados en los discursos).

Entre los módulos incluidos en los componentes contextuales, incluye un motor de fusión multimodal, módulos de procesamiento discursivo para la resolución de diversos tipos de fenómenos referenciales y elípticos y también módulos para la gestión de la toma de turno, la identificación del destinatario y la generación de realimentación.

El sistema FADE modela el estado de la interacción y el de la palabra y dispara, en función de ambos, nuevos turnos del sistema. Para ello aplica reglas relacionadas en el estado de los turnos del resto de participantes. Por ejemplo, si la palabra está vacante y otro participante

comienza a enunciar el sistema considera que dicho participante pasa a estar en posesión de la palabra.

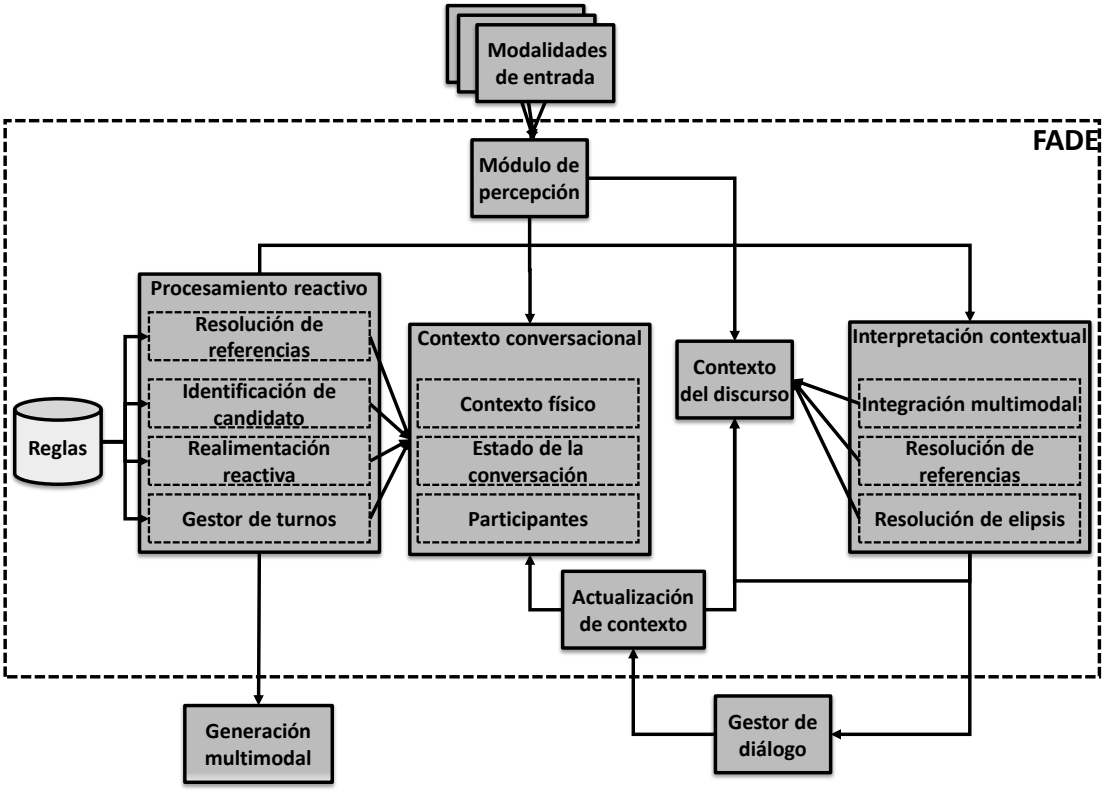


Figura 6: Arquitectura Fade [139]

La virtudes adicionales de esta propuesta sobre el sistema Ymir son la de tener una orientación multimodal y la de soportar *interpretación incremental* de las enunciaciones del resto de participantes. Esto último permite actualizar dinámicamente el contexto a medida que la enunciación va siendo interpretada (y no sólo tras su completa interpretación). La incorporación de la interpretación incremental permite en este sistema generar realimentación simultánea a las contribuciones de sus interlocutores [120].

En cuanto a las carencias, al igual que ocurría con el sistema anterior, FADE, no está capacitado para generar incrementalmente la enunciación del sistema, por lo que no tolera rectificaciones, reformulaciones o interrupciones. Todos los cambios producidos en el contexto afectarán sólo a las enunciaciones futuras del sistema y en ningún caso a la que esté en curso. Por otro lado, y a pesar de ser capaz de tomar el turno incluso de forma solapada para producir realimentación, en lo referente a sus intervenciones sigue teniendo un comportamiento pasivo, limitándose a estimar cuando el resto de participantes esperan del sistema que tome el turno.

3.4.1.3 VM-GEN

VM-GEN [55] (parte del proyecto VERBMOVIL [86]) es un componente de *generación incremental* de enunciaciones habladas. Su sistema está articulado en interfaz de entrada, formulador de frases, componente de linearización e interfaz de salida [Figura 7].

Se basa en la definición propuesta por Kempen y Hoenkamp [103] del término *generación incremental*. Según esta definición, las personas pueden comenzar a hablar sin tener completamente formulado aquello que quieren decir y que, sólo mientras comienzan a hablar, van refinando en contenido y forma. Además, la definición especifica que en una generación incremental el habla debe ser articulada y fluida, incluso aun sin que la enunciación sea perfecta o completa. Según este planteamiento, primero se realiza una formalización preliminar, y queda abierta la posibilidad de que el generador complete y refine la enunciación prevista, obteniendo la mejor expresión posible desde un punto de vista gramatical.

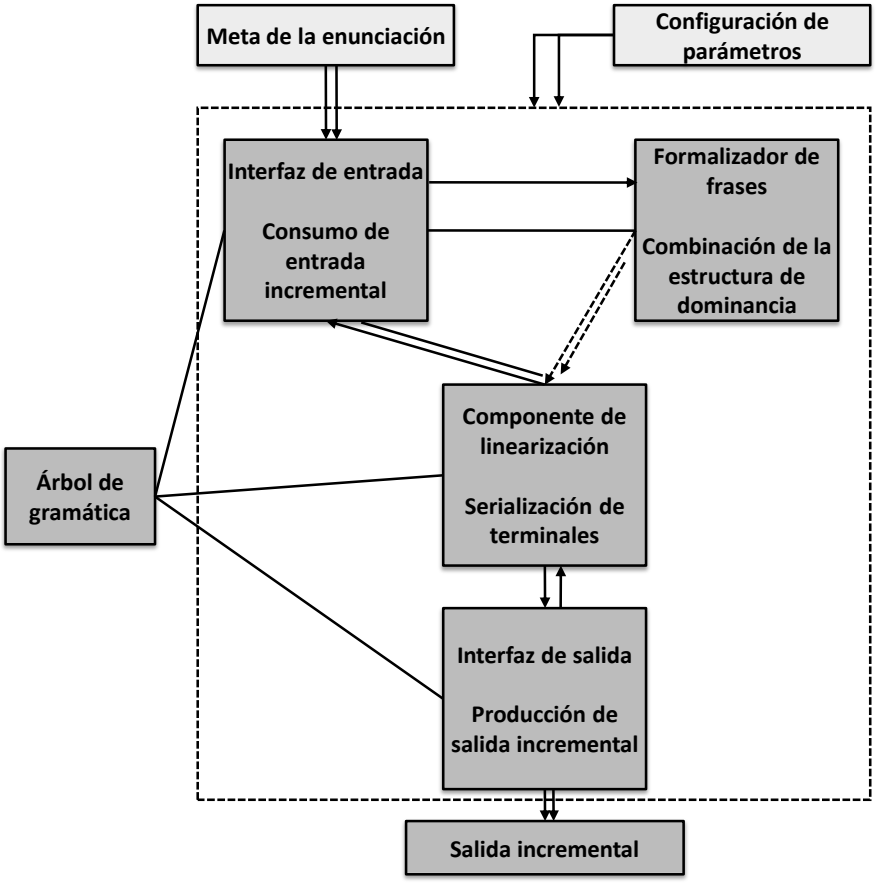


Figura 7: Arquitectura de VM-GEN [55]

Para desarrollar la generación incremental VM-GEN) aplica los siguientes principios:

1. Cada nuevo incremento recibido a la entrada (durante la interpretación de las enunciaciones del usuario) dispara un nuevo proceso de generación, de forma que:

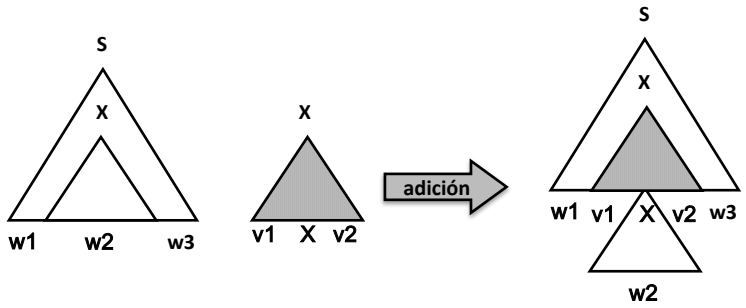
    a. El procesamiento de la generación comenzará antes de que la interpretación esté completa.

    b. Quepa la posibilidad de que sean obtenidas generaciones preliminares, incluso aunque el proceso de generación haya terminado.

    c. Quepa la posibilidad de que sean obtenidas generaciones preliminares, incluso aunque el proceso de interpretación haya terminado.

2. Los incrementos de la salida deben resultar del procesamiento de algo pasado.

La clave de la generación incremental son los Árboles de Gramáticas Contiguas (TAG) [106, pp. 20-24] que el sistema aplica para realizar la representación sintáctica. Estos árboles permiten representar un predicado y todos sus argumentos como un árbol, y puede representar restricciones en cuanto a las subcagorizaciones de la información y a la necesidad de coincidencia entre el contenido de determinados nodos. En definitiva, los nodos en estos árboles representan propiedades estructuradas, con lo que es posible realizar unificaciones [Ejemplo 7] entre distintos árboles (distintos predicados).



Ejemplo 7: Ejemplo de unificación de Árboles de Gramáticas Contiguas (TAG) [106, pp. 20-24]

Esto permite adaptar en el formulador de frases la enunciación del sistema durante la marcha, a medida que evoluciona la interpretación adquirida por el interfaz de entrada. El componente de linearización se encarga de expresar de forma fluida a lo largo del tiempo los árboles sintácticos generados, tratando cuestiones como la rectificación ante cambios en el árbol. Finalmente, la interfaz de salida expresa la contribución previamente linearizada y monitoriza el estado de expresión alcanzado.

En definitiva, VM-GEN ofrece soluciones a determinados aspectos relacionados con la generación incremental (capacidad para modificar dinámicamente la enunciación del sistema y gestión de su continuidad), pero no trata el resto de aspectos que componen el problema de una *toma de turno avanzada* en la interacción.

3.4.2 DECOP

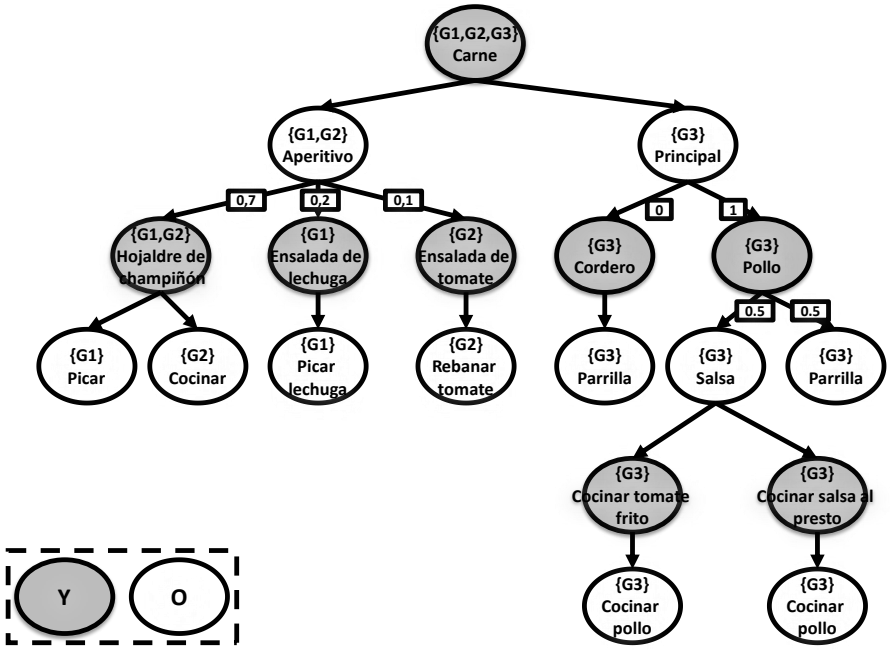
DECOP [101] consiste en un modelo de toma de decisiones colaborativo en el que el sistema debe decidir cuándo desarrollar una meta o no, considerando tanto los intereses de los interlocutores humanos como los suyos propios. Los participantes de la interacción colaboran entre ellos en una acción combinada con el objetivo de satisfacer sus propias metas. Bajo este planteamiento, existen dos posibles decisiones a considerar: la de desarrollar o no una meta y la de aceptar o no su desarrollo (en función de que se sea el participante del que parte la iniciativa o el que la recibe). Ambas acciones tienen asociado un coste y un beneficio mutuo, y los participantes deben estimar ambos para realizar su toma de decisión.

DEPOC representa las metas de la interacción en lo que denomina Árboles de Receta Probabilística (PRTs, del inglés Probabilistic Recipe Trees). Con ellos, cada meta se descompone en otras más sencillas, define a qué participantes implica y determina la probabilidad de ser realizada frente a otras. El Ejemplo 8 muestra el PRT asociado a la meta "preparar la cena". Las hojas representan las metas simples que la componen (por ejemplo, "cortar champiñón") y los nodos intermedios la forma en que se combinan en otras más complejas ("fritura de champiñón" está compuesto por "cortar champiñón" y "cocinar").

Los árboles permiten calcular la probabilidad de que una meta sea realizada frente a otras, el coste de hacerlo para cada agente y estimar si tiene o no un beneficio mutuo. Tanto el coste de desarrollar una meta como su beneficio mutuo son dependientes del contexto, por lo que al progresar la interacción cambian. De esta forma, en cada instante es posible decidir si desarrollar o no una meta (introducirla o aceptarla) en función de si el beneficio mutuo es mayor que su coste.

Aunque DECOP permite modelar la utilidad de que un participante (el sistema) desarrolle las metas de la interacción, dicho modelo se aplica sobre arquitecturas restringidas a una toma de turno por *ciclo de interacción* [Apartado 2.1]. Por ello, a efectos prácticos, sólo permite determinar si el sistema, durante su turno, debe o no desarrollar cada una de las metas de la interacción. Aunque la propuesta es potencialmente aplicable a la gestión de metas, queda pendiente independizarla, como proceso, del resto de los procesos de la *interacción natural*, para que dicha decisión pudiera afectar a un verdadero reparto de turnos. Sin conocimiento

sobre el estado de los turnos, de la posesión de la palabra y de los participantes que son candidatos a tomarla, su aplicación a la decisión de toma de turno queda limitada. Junto a ello, sin procesamiento incremental ni gestión de la continuidad, quedan fuera de las posibilidades del sistema el adecuar su contribución en curso a los cambios en las circunstancias en las que se produce. En definitiva, la propuesta no puede ser aplicada a la gestión de los solapamientos, interrupciones, reformulaciones, rectificaciones o auto interrupciones de contribuciones.



Ejemplo 8: PRT para la acción "preparar la cena" [101].

3.5 *METODOLOGÍAS DE EVALUACIÓN DE SISTEMAS DE INTERACCIÓN NATURAL*

Varios autores han propuesto metodologías de evaluación de Sistemas de Interacción Natural. De hecho, existen casi tantas metodologías de evaluación como Sistemas de Interacción Natural han sido desarrollados hasta la fecha. Algunas como ATIS [137] se limitan a medir del funcionamiento del sistema completo, sin permitir la extracción de conclusiones particulares para cada uno de los componentes de la arquitectura. Evalda [123] sí permite una evaluación independiente de algunos de los componentes de la arquitectura, aunque no incluye entre ellos ninguno de los responsables de la gestión del diálogo y la toma de turno. Churcher et al. [30] introducen una metodología para la evaluación de la gestión del diálogo. Para ello parte de una medida de las capacidades del sistema en aspectos como la resolución de anáforas y elipsis,

estrategias interactivas aplicadas, etc., cuyo peso en la naturalidad de la interacción es determinada a partir de la opinión de expertos.

Estas son metodologías orientadas a la evaluación individual de sistemas específicos, por lo que cada una de ellas presta atención a aspectos muy diferentes y se hace difícil la comparación de sus resultados. Este problema ha dado lugar a otras metodologías como EAGLES [107], que aplica la normativa ISO 9000 para definir unos entornos, criterios y métricas de evaluación que permiten confrontar los resultados obtenidos para distintos sistemas.

Todas ellas abordan la *evaluación objetiva del funcionamiento del sistema* y de sus subcomponentes, y se basan en parámetros *cuantitativos* (productos concretos del funcionamiento del sistema) y *cualitativos* (estimaciones y juicios sobre las propiedades del sistema a partir de la aplicación de estándares, reglas y conocimiento experto). Algunos de los parámetros cuantitativos más comúnmente considerados son: tiempo de resolución de las metas; número de metas resueltas; número de turnos desarrollados; o duración de la interacción. Por su parte, la evaluación cualitativa considera cuestiones como: porcentaje de metas de usuario identificadas correctamente; corrección, relevancia y suficiencia informativa de los contenidos de los turnos del sistema; adecuación de la iniciativa del diálogo; o detección, clasificación, diagnóstico y reparación de problemas de interpretación, lingüísticos, de coherencia discursiva, etc.

Tanto Calle et al. [20] como Bernsen y Dybkjaer [9] ponen de manifiesto que, si de lo que se trata es de medir el grado de naturalidad en la interacción, la *usabilidad* debe toma un papel muy relevante en la evaluación. Partiendo de esta idea Bernsen y Dybkjaer proponen la metodología DISC, que incorpora medidas de la satisfacción del usuario basándose en los estudios de Grice [78]. De esta forma, la usabilidad se mide en términos de la capacidad manifestada por el sistema para cooperar con el usuario durante la interacción. Para ello considera parámetros como: la falta o exceso de la información proporcionada por el sistema; problemas en su veracidad o relevancia; falta de claridad o ambigüedad en sus intervenciones; exceso de duración o problemas de orden es sus intervenciones; no consideración del contexto de la interacción; no gestión de errores; etc.

La evaluación de la usabilidad es, a menudo, realizada sobre *parámetros subjetivos*. Estos son obtenidos a través de cuestionarios y entrevistas, lo que, según algunos estudios [51], conlleva una serie de problemas asociados. En primer lugar, los usuarios de prueba no son usuarios reales y, aun cuando éstos puedan constituir un grupo suficientemente representativo, el sólo hecho de conocer que se participa en un proceso de evaluación y cuál es el objetivo de la evaluación, generan unas expectativas sobre el resultado y tienen una clara influencia en sus

valoraciones. Por otro lado, a medida que los usuarios realizan pruebas con el sistema se familiarizan con la forma en la que deben expresarse y guiar la interacción, por lo que la evaluación subjetiva puede mejorar a lo largo del tiempo, sin que el sistema haya realmente mejorado. Por último, lo que cada usuario entiende por naturalidad puede ser muy distinto y, para cada uno de ellos, la importancia relativa de los aspectos que la componen puede ser diferente (por lo que los datos proporcionados por distintos usuarios no son siempre comparables). Dado que la evaluación de la usabilidad es imprescindible para medir el grado de naturalidad de la interacción, Walker et al. [180] proponen la metodología PARADISE (posteriormente extendida a sistemas multimodales bajo el nombre PROMISE [7]) que pretende independizar la evaluación de la usabilidad de factores subjetivos. Esta metodología toma como principal indicador la satisfacción del usuario y proponen estimarla a partir de medidas estrictamente objetivas (tasa de éxito en la resolución de tareas, eficiencia del sistema, calidad y coste del diálogo, etc.). No obstante, no existen resultados concluyentes que permitan afirmar que es posible independizar la evaluación de la usabilidad de la valoración subjetiva del usuario.

Por lo general las habilidades de toma de turno han sido uno de los aspectos menos atendidos en la evaluación de los sistemas de interacción. Describir una metodología de evaluación más adaptada a estas necesidades hace preciso identificar el conjunto de rasgos que caracterizan la toma de turno humana. Algunos autores han tratado de describir el conjunto de marcadores léxicos, sintácticos, prosódicos y gestuales que intervienen en ella. Estudios como los de Duncan [45;44;46;47] y Duncan y Fiske [48] conjeturan que los hablantes manifiestan al final de sus turnos un conjunto de señales complejas compuestas de rasgos entonacionales, prosódicos, gesticulares, cambios en el timbre y de expresiones estereotipadas. Mushin et al. [130] estudian la prosodia de los asentimientos. Koiso et al. [109] identifica dependencias entre la realimentación y la sintaxis y prosodia. Ward and Tsukahara [186] describen la prosodia de las invitaciones mostradas por el hablante a sus oyentes para que ofrezcan realimentación. Tanto Ford y Thompson [59] como Wennerstrom y Siegel [190] analizan la correlación que existe entre la completitud gramatical y entonacional con los cambios de hablante. Schaffer [152] o Cutler y Pearson [42] presentan estudios en los que se comparan las diferencias existentes entre los marcadores de toma de turno que se expresan en las interacciones cara a cara y los que se expresan en otro tipo de interacciones. Otros autores, como Cathcart et al. [25] y Poppe et al. [141], han tratado de aplicar los resultados de estos estudios a la predicción de la realimentación y a la detección de los límites de las contribuciones de los participantes [58;57;52;157;4;144].

En general, estos estudios caracterizan la forma en que los marcadores de toma de turno son expresados por los participantes a través del lenguaje y del paralenguaje (prosodia, entonación, gestos, etc.) o tratan de simplificar a una relación probabilística de causa-efecto la

expresión de estos marcadores con los cambios en la posesión de la palabra [76]. Ninguno de ellos pretende analizar cuáles son los mecanismos del uso del lenguaje que desencadenan finalmente eventos en la toma de turno a partir de su ocurrencia (solicitudes o cesiones de palabra; decisiones de toma, mantenimiento de turno o su liberación; decisión de producción de turnos primarios o secundarios; etc.). Desde el punto de vista de la pragmática, no es tan relevante la descripción de las posibles realizaciones de estos marcadores a través de los distintos códigos y modalidades humanos (su expresión, propiamente dicha), sino la descripción de los movimientos que estos producen en la interacción, y cómo dependiendo de las circunstancias sociolingüísticas y del estado en que se encuentran las metas de los participantes repercuten en la toma de turno. En relación a este problema, Jurafsky et al. [99] describen que las realizaciones léxicas de los actos de diálogo son determinantes en la identificación de la realimentación. Por otro lado, Yngve [196], Duncan [45], Kendon [105], Schegloff [153], Jefferson [93] y Novick y Sutton [132] proponen categorizaciones de la realimentación lingüística en los diálogos orientados a tarea, los cuales están basados en el contexto estructural de la interacción. Besser y Alexandersson [12] han tratado el problema de la identificación de los posibles tipos de disfluencia presentes en la interacción. Sacks et al. [149] describen el conjunto de acuerdos tácitos que regulan el traspaso de la palabra en la interacción humana. Finalmente, Clark [32] matiza cuáles son el conjunto de factores que podrían alterar este sistema de toma de turno.

Entre las metodologías de evaluación que consideran aspectos relacionados con la toma de turno, destaca la propuesta por Kammar el al. sobre su sistema DECOP [102], que estima la oportunidad de interrumpir en la interacción en base al cálculo de la relación beneficio-coste de hacerlo. Para su evaluación proponen un escenario de pruebas en el que usuario y sistema deben colaborar para resolver una tarea y confrontan su propuesta con otras estrategias de interrupción. Consideran en la evaluación parámetros como el nivel de éxito alcanzado en la resolución de la tarea y la tasa de aceptación de la interrupción por los usuarios. No obstante, su propuesta se restringe a una estimación de la oportunidad de la interrupción, sin prestar atención al resto de aspectos que toman parte en el correcto desarrollo de la toma de turno: el procesamiento incremental; la gestión de la continuidad en las contribuciones de los participantes; la representación del estado de los turnos de los participantes; la representación del hablante actual; y la representación de los candidatos a tomar la palabra.

3.6 *CONCLUSIONES*

La toma de turno es, hasta el momento, uno de los aspectos menos tratados en el desarrollo de los Sistemas de Interacción Natural. Los más adecuados para abordar una toma de turno similar a la de la *interacción humana* (una *toma de turno avanzada*) son los sistemas *discursivos*, *multimodales*, de *iniciativa mixta* y de *acción combinada* [3.1.5]. Junto a esto, se requieren soluciones que aborden la *interacción natural* desde la independencia de procesos y un desarrollo temporal más realista que el basado en el *ciclo de interacción* [Apartado 2.1].

En algunos casos [Tabla 6], se atienden aspectos como la *interpretación* y *generación incremental* (FADE [Apartados 3.4.1.2]), incluso abordando aspectos de la *gestión de la continuidad* de la contribución ([VM-GEN [Apartado 3.4.1.3]). También existen modelos que permiten detectar marcadores relacionados con la toma de turno, y estimar cuándo se espera del sistema que tome la palabra (FADE e Ymir [Apartado 3.4.1.1]). Por su parte, la propuesta de metodología de toma de decisiones colaborativas DECOP [Apartado 3.4.2], propone un mecanismo de *gestión de metas* que permite estimar cual es el coste y los beneficios de desarrollar una meta en la interacción, considerando el beneficio colaborativo de todos los participantes.

Tabla 6: Tabla comparativa de las habilidades de *toma de turno avanzada* soportadas por los distintos sistemas

Sistema	Independencia de Procesos	Gestión de Continuidad	Gestión de Metas	Marcadores de Toma de Turno	Estado de Turnos, Palabra y Candidatos	Decisión de Toma de Turno
Ymir [169]	NO INCREMENTAL	NO	NO	Sí	Sí	PASIVA
FADE [139]	Incremental	NO	NO	Sí	Sí	PASIVA
VM-GEN [55]	Incremental	SÍ*	NO	NO	NO	NO
DECOP [101]	NO	NO	Sí	NO	NO	NO

(*) Sin detección de marcadores producidos como alteración de la continuidad temporal de la contribución.

Aunque con esto quedan abordados gran parte de los problemas que entraña una toma de turno avanzada, no existe una solución completa a todos ellos y, en cualquier caso, quedan aún muchas cuestiones por ser tratadas.

De los sistemas que abordan la independencia de procesos, sólo se considera el procesamiento incremental en dos de ellos, y sólo en uno se incluye la gestión de la continuidad. Incluso en este caso, no se considera la forma en la que los procesos de reinterpretación afectan

al estado en el que se encuentra la interacción y el resto del conocimiento sociolingüístico involucrado en ella. Algunos ejemplos soportan generación incremental de la enunciación cuando los cambios se restringen a modificaciones sobre el predicado en curso, pero no cuando afecta a alteraciones más profundas de su enunciación, como son los casos de auto interrupción (cambios de tema espontáneos) o cancelación de metas ("*Ah, bueno. Entonces nada*"). Tampoco son tratados los marcadores relacionados con la toma de turno que los participantes expresan como alteraciones de la continuidad temporal de sus contribuciones [165]. La gestión de metas también ha sido aplicada, pero sin ser considerada en combinación con otros aspectos de la toma de turno, la mejora que aporta sobre la naturalidad de la interacción queda muy limitada. En lo que respecta a la *gestión de la toma de turno*, existen sistemas capaces de detectar marcadores expresados de forma explícita de forma verbal o a través de gestos. También los hay que modelan determinados aspectos del estado de posesión de la palabra e identifican los participantes que se perfilan como candidatos a tomarla. Además, los sistemas que representan esta información, implementan estrategias de decisión de toma de turno. Sin embargo, en todos estos casos la gestión de toma de turno queda reducida a estimar cuándo los interlocutores esperan del sistema que participe, tomando un papel estrictamente pasivo. No se contempla la propia iniciativa del sistema en lo referente a la toma de turno (en función del estado de las propias metas del sistema, el estado de la interacción, la situación, las metas emocionales o el resto de conocimiento sociolingüístico involucrado). Quedan fuera de las capacidades de los Sistemas de Interacción Natural acciones tan comunes en la interacción humana como: contribuir solapadamente (por ejemplo para ofrecer realimentación simultánea); interrumpir o solicitar la palabra (por el interés general de los participantes o por un aumento en la urgencia de las metas propias del sistema); ceder la palabra (cuando sea favorable para el interés general); o desarrollar estrategias orientadas a retenerla (cuando otros participantes interrumpen injustificadamente); etc.

Por todo ello, se requieren nuevas arquitecturas de Sistemas de Interacción Natural que permitan hacer del desarrollo temporal de la interacción un proceso menos mecánico y artificial, que integren todas las habilidades descritas y en la que el sistema no se limite a tomar una actitud pasiva en la toma de turno.

Finalmente, en lo que respecta a la evaluación de los Sistemas de Interacción Natural, no existen metodologías que describan de forma completa mecanismos y métricas adecuados para la evaluación de estrategias de toma de turno. Del mismo modo, estas metodologías están diseñadas para proporcionar medidas relativas de la naturalidad de la interacción de unos sistemas frente a otros, pero no ofrecen información sobre cómo de próxima es esta naturalidad a la alcanzada por propia interacción humana. De esta forma, se precisan metodologías de

evaluación que permitan analizar el impacto que tiene el procesamiento incremental y la decisión de toma de turno sobre la naturalidad de la interacción. Metodologías que no estén limitadas a un análisis estadístico de la frecuencia con la que se producen marcadores concretos de toma de turno y que consideren aspectos de carácter pragmático que permitan medir la repercusión que una mejor alternancia de turnos tiene sobre la satisfacción del usuario y la resolución colaborativa de las metas de la interacción. En estas metodologías deberán ser considerados tanto aspectos técnicos y objetivos, como la propia medida de la usabilidad (la cual dependerá de las valoraciones subjetivas aportadas por los usuarios). Las valoraciones subjetivas serán obtenidas sin que las expectativas de los usuarios sobre la evaluación puedan condicionar su resultado, y sin que puedan hacerlo tampoco sus diferencias de entrenamiento con cada una de las configuraciones evaluadas. Del mismo modo, deberán aplicarse métricas que permitan hacer comparables las valoraciones ofrecidas por los distintos usuarios (dado que cada una de ellas responde a lo que cada usuario entiende por naturalidad).

Capítulo 4 MARCO DE LA PROPUESTA

Esta tesis se enmarca dentro de los trabajos del Grupo de Bases de Datos Avanzadas de la Universidad Carlos III de Madrid, y consiste en un Sistema de Interacción Natural con gestión del diálogo de acción combinada y una gestión avanzada de la toma de turno para mejorar el desarrollo temporal de la interacción. El Sistema de Interacción Natural que aquí se presenta está concebido para ser desplegado sobre la Plataforma Ecosistema, un entorno Multi Agente de Pizarra Compartida [Apartado 3.3.2] que hace posible desarrollar de forma concurrente los distintos procesos que tienen lugar en la *interacción natural*.

A lo largo del presente apartado se describen, tanto la arquitectura de Sistema de Interacción Natural sobre la que se construye la propuesta, como la Plataforma Multi Agente de Pizarra Compartida Ecosistema [Apartado 4.3].

4.1 *ARQUITECTURA COGNITIVA PARA UNA TOMA DE TURNO AVANZADA*

La interacción desarrollada entre las personas, la *interacción humana*, requiere el manejo de gran cantidad de conocimiento y la puesta en juego de numerosas habilidades. Los sistemas que pretenden reproducir este comportamiento interactivo deben, por tanto, implementar componentes que permitan reproducir dichas habilidades y representar la mayor cantidad posible de todo ese conocimiento. Cuantas más habilidades interactivas sean implementadas y más completo sea el conocimiento manejado por el sistema, más natural será la interacción que éste puede desarrollar. De esta forma, el objetivo del presente apartado será estructurar en un conjunto de componentes todo el conocimiento y las funciones de los que depende el tratamiento de la interacción. Para ello, se toman como punto de partida los trabajos

de Calle [21], que describen una arquitectura de Sistema de Interacción Natural *discursiva*, *multimodal*, de *iniciativa mixta* y de *acción combinada* [Apartado 3.1]. Ésta consta de los siguientes modelos de conocimiento [Figura 8]:

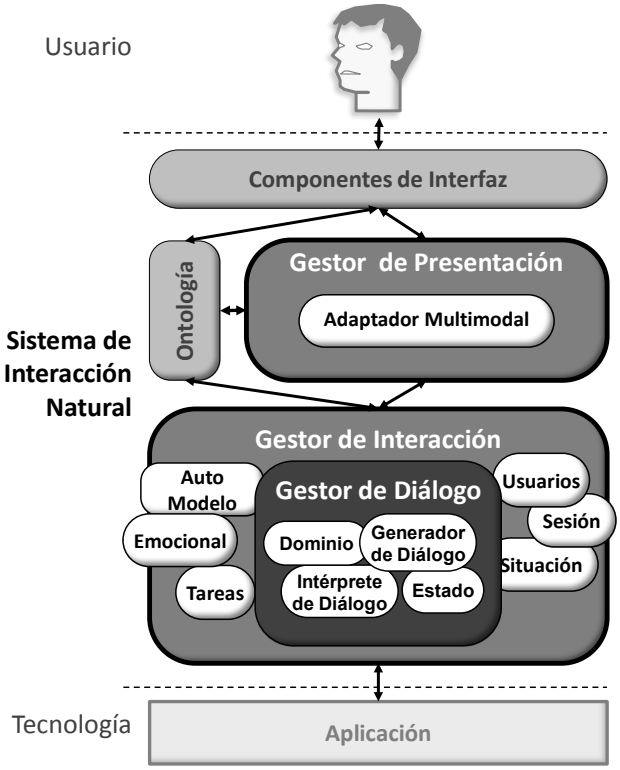


Figura 8. Arquitectura cognitiva para la *interacción natural*

- *Componentes de Interfaz*: Adquieren y sintetizan contribuciones en modalidades y códigos naturales al usuario.
- *Ontología*: Permite hacer comparables los conceptos que maneja el sistema con los que maneja el usuario, y los que manejan los distintos modelos de conocimiento entre ellos.
- *Gestor de Interacción*: Este componente se encarga de las funciones de gestión del diálogo (representación del dominio de interacción, del estado de la interacción, interpretación de diálogo de las contribuciones del usuario y generación de diálogo de las del sistema). También representa el resto de conocimiento sociolingüístico asociado a la interacción, que es actualizado por el gestor de diálogo durante los procesos de generación e interpretación de diálogo. Dicho conocimiento ser articula en: Modelo de Sesión, Modelo de Usuario, Modelo de Situación, Modelo Emocional y Auto Modelo. Finalmente, el Componente de Interacción es el puente entre el usuario y la tecnología a la

que el usuario desea acceder, para lo cual representa el conocimiento operativo de la interacción en el denominado Modelo de Tareas.

- *Gestor de Presentación*: Es el componente que tradicionalmente soporta las funciones de fusión y fisión multimodal (transformación de las expresiones multimodales del usuario en un flujo de datos único y adaptación de las contribuciones del sistema a modalidades y códigos naturales, respectivamente).

A lo largo del presente apartado serán descritos los Componentes de Interfaz [4.1.1], la Ontología [4.1.2] y el Gestor de Presentación [4.1.3]. El Gestor de Interacción, por su extensión, será tratado en un apartado posterior [Apartado 4.1.3].

4.1.1 Componentes de Interfaz

La interfaz atiende a la adquisición de las expresiones naturales enunciadas por el usuario y a su representación en estructuras semánticas formalizadas de forma que puedan ser procesadas por el sistema. También a la síntesis de las contribuciones generadas por el sistema en modalidades y códigos adecuados para el usuario. Dichas tareas son realizadas, respectivamente, por los subcomponentes de interfaz de entrada y de salida.

4.1.1.1 Interfaz de Entrada

La Interfaz de Entrada es el conjunto de componentes de un Sistema de Interacción Natural que posibilitan la adquisición de las expresiones de los usuarios en forma de voz, texto o cualquiera de las modalidades que pudiesen ser soportadas en la interacción (funciones físicas). Además, extrae de ellas su contenido semántico, representándolas en formatos procesables por el componente de Adaptación Multimodal [4.1.3.1], que las recibirá una vez tratadas por la interfaz (funciones lógicas).

Para cada una de las posibles modalidades de entrada, el sistema deberá implementar un par de componentes físico-lógicos adecuados para las labores de procesamiento de las expresiones recibidas en dicha modalidad. De esta forma, los componentes físicos son [Figura 9] aquellos que reciben directamente de los sensores (micrófonos, teclados, cámaras, ratones, pantallas táctiles, etc.) las señales físicas emitidas por los usuarios. Los lógicos son aquellos que extraen de ellas las características correspondientes a la modalidad que procesan, y prescinden del resto de información contenida en las señales originales. Así, los dispositivos de reconocimiento de voz extraerán de la señal de audio la intervención realizada por el usuario (habitualmente en forma de texto, con anotaciones sobre entonación, pausas, etc.); los de reconocimiento de gestos detectan, dentro de las señales de vídeo, los gestos realizados por el

usuario y los codifican en forma de símbolos (dentro de una representación simbólica adecuada al dominio de gestos posibles de dicha modalidad [104; 135]).

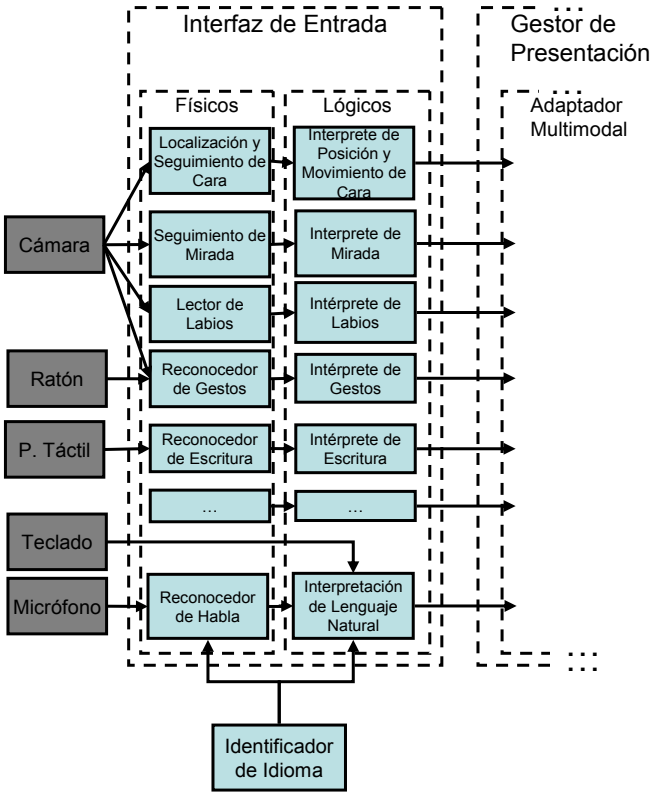


Figura 9. Arquitectura Cognitiva de la Interfaz de Entrada

Cuando el dispositivo es multilingüe y la selección del idioma es automática, se requieren además componentes capaces de determinar el idioma en el que se ha producido la intervención (Componentes de Identificación de Idioma). Estos componentes basan la decisión en análisis acústicos y lingüísticos sobre la señal recibida. Su objetivo es encontrar el idioma al que más probablemente pertenece una determinada contribución producida a la entrada. Para ello, además de la propia contribución, a menudo consideran también otras contribuciones recibidas anteriormente. Cuando la arquitectura incluye un Modelo de Situación [Apartado 4.2.3.3], las funciones del los Componentes de Identificación de Idioma serán cubiertas por él, como parte del aspecto semiótico.

4.1.1.2 Interfaz de Salida

La interfaz de salida está constituida por el conjunto de componentes que posibilitan la expresión coordinada de las contribuciones que generó el Gestor de Diálogo [Apartado 4.2.1], y

cuya presentación a través de diferentes modalidades y lenguas fue diseñada por el componente de Adaptación Multimodal [Apartado 4.1.3.1].

Los componentes de salida más comunes son los sintetizadores de voz (por ejemplo Festival [35]) y los avatares gráficos (que presentan de forma combinada gestos faciales, corporales, sincronización labial, situación y señalamiento, entre otros [23]). Entre ellos se pueden encontrar también representaciones de figuras, mapas, diagramas, texto, etc.

Al igual que los componentes de la Interfaz de Entrada, los de la Interfaz de Salida se dividen entre componentes lógicos y físicos [Figura 10]. En la interfaz de salida, cada modalidad necesita tanto componentes lógicos como físicos, y en ellos se realizan las tareas complementarias a las realizadas en la Interfaz de Entrada. El componente lógico de una determinada modalidad recibe del Adaptador Multimodal el flujo de datos que deberá expresar su correspondiente componente físico a través de dicha modalidad. Así, el componente Generador de Lenguaje Natural recibirá una secuencia de símbolos a partir de la cual deberá producir una expresión textual (con posibles anotaciones) que serán remitidas al Sintetizador de Habla. Otros, como el Generador de Mirada, transformarán los símbolos recibidos en puntos del espacio y objetos concretos sobre los que el avatar deba fijar su mirada, etc.

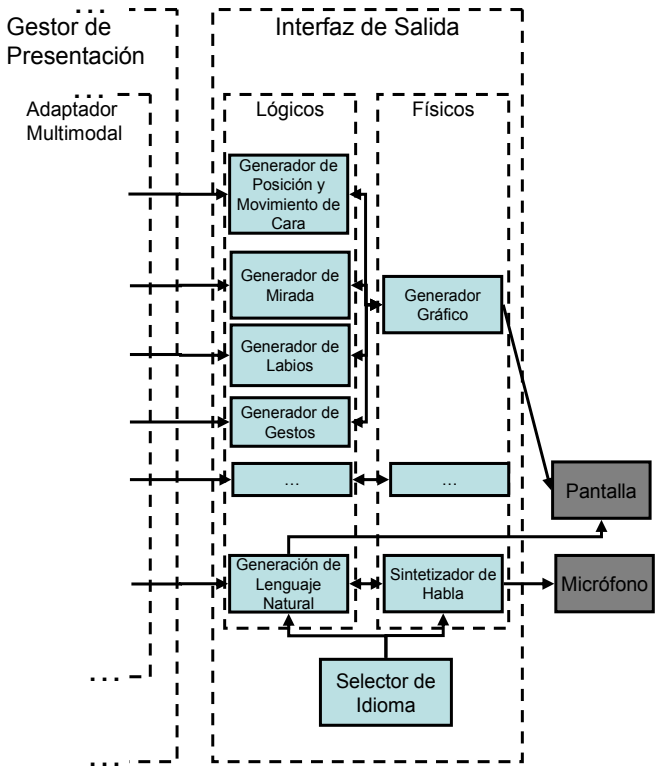


Figura 10. Arquitectura Cognitiva de la Interfaz de Salida

Las expresiones correspondientes a las distintas modalidades deberán producirse de forma coordinada, y aquellas modalidades que resulten complementarias (como es el caso del habla y el movimiento de los labios) requieren una perfecta sincronización entre ellas. Por ello, resulta fundamental que los componentes de salida de la interfaz ajusten con precisión la síntesis de sus expresiones al diseño de la contribución que realizó previamente el Adaptador Multimodal. Cuando el sistema soporta múltiples idiomas, se requieren componentes específicos de generación de Lenguaje Natural para cada uno de ellos. En el caso de los sintetizadores de voz, pueden ser específicos para cada lengua o multilingües como Festival.

4.1.2 Ontología

La Ontología es el componente que hace comparables los conceptos que maneja el Sistema de Interacción con los que manejan el resto de participantes. A la vez, hace comparables entre ellos los conceptos que manejan internamente los distintos componentes de la arquitectura [82].

Los conceptos son formas de entendimiento de realidades o abstracciones (*significados*) a las que se les asocian representaciones simbólicas específicas (*significantes*). La representación simbólica de significados hace posible definir sus propiedades (y los dominios de éstas); las relaciones que mantienen entre ellos (instancia, generalización, composición, etc.); y las restricciones que presentan en sus propiedades y relaciones. Los conceptos pueden organizarse en familias semánticas de acuerdo a determinadas propiedades (*semas*). En función del dominio de valores posibles para un determinado sema, la familia semántica puede estar compuesta por un conjunto no continuo de conceptos (por ejemplo, la familia de colores), o admitir gradaciones (si, por ejemplo, se toma por sema la luminosidad).

En cuanto a la utilidad de la ontología para hacer comparables los conceptos manejados por el sistema y por el resto de los participantes, el sistema debe de ser capaz de reconocer y expresar los distintos conceptos contenidos en el dominio de interacción, esto es el área de experiencia común a los participantes dentro del cual puede discurrir la interacción. En la medida en la que dicha ontología recoja también los distintos sinónimos que permiten referirse a un mismo concepto, la probabilidad de éxito de la interacción será mayor, pero esto no resulta una tarea sencilla debido a que, más a allá de lo elevado del número de conceptos que pueda requerir la interacción en un determinado dominio, las ontologías manejadas por los distintos participantes son también distintas. Más aún, debido a la naturaleza personal y en gran medida subjetiva de las ontologías manejadas por los humanos, en muchos casos la relación entre conceptos y significantes divergen de forma muy notable en las ontologías de los distintos

participantes. Además, a menudo los individuos modifican sus ontologías dinámicamente como resultado de su experiencia.

De esta forma, aunque es imprescindible que las ontologías manejadas por los distintos participantes sean comparables para hacer posible la interacción, en la práctica no siempre se podrá encontrar una correspondencia exacta entre los símbolos emitidos por un participante y los conceptos que referencian en las ontologías del resto. En su lugar, para cada símbolo percibido, el sistema deberá escoger entre un conjunto de posibles conceptos en base a preferencias sobre el dominio, la situación, el usuario, etc. Aparece entonces la necesidad de medir la certeza en la elección realizada, y esta certeza afectará a los cambios que se produzcan en el estado de la interacción y a las enunciaciones que realizará el sistema en consecuencia. Cuando la certeza es muy baja (o similar para varios conceptos), pueden aplicarse estrategias interactivas para resolver la incertidumbre (interrupciones, confirmaciones explicitas o implícitas, etc.) por medio de la introducción de nuevas metas discursivas en la interacción a través del Gestor de Diálogo [Apartado 4.2.1]. Por último, gran parte de los conceptos representados serán comunes a distintos dominios. En la medida de lo posible, debe cuidarse el diseño de las ontologías desarrolladas para aumentar su robustez y permitir la reutilización de los conceptos implementados en otros dominios.

Pero la importancia de la Ontología no sólo radica en permitir un entendimiento entre el usuario y el sistema. Dada la diversidad de componentes que integran la arquitectura, y los números intercambios que se establecen entre ellos a lo largo de la interacción, es también necesaria una correspondencia común entre los símbolos y significados que comparten entre ellos. De esta forma, las solicitudes y resoluciones de servicio llevadas a cabo por los componentes, deberán acogerse a su ontología común. En los casos en los que esto no sea posible, por tratarse de componentes externos relacionados con aspectos operativos (las aplicaciones que deben ser invocadas para satisfacer las necesidades del usuario), se requerirá la implementación de capas de persistencia para adaptar las operaciones específicas que ofrecen estos componentes a la ontología propia del sistema. Esta función corre a cargo del Gestor de Tareas, en el Gestor de Diálogo, dentro del Componente de Interacción [Apartado 4.2.2].

4.1.3 Gestor de Presentación

En un escenario en que los participantes intervienen en orden, de uno en uno, y durante el tiempo que ellos, como hablantes, consideran oportuno (según el *ciclo de interacción* [Apartado 2.1]), la interacción puede ser tratada por el Gestor de Presentación como una sucesión ordenada de fases, donde las contribuciones del usuario son interpretadas tras haber

sido recibidas por completo; el sistema conmuta de rol de oyente al de hablante tras la completa interpretación de la contribución del usuario; y, sólo después de haber sido expresada la contribución que el sistema genera como respuesta, vuelve a tomar el rol de oyente. Bajo este planteamiento, la presentación se restringe a las funciones de adaptación multimodal. Es decir, a la fusión (combinación) de las expresiones recibidas del usuario (a través de diferentes modalidades y lenguas) en un único mensaje fácilmente procesable por el sistema, y a la fisión (descomposición) del mensaje generado por el sistema en expresiones de distintas modalidades y lenguas naturales al usuario.

4.1.3.1 Adaptador Multimodal

En lo referente a la adaptación multimodal, pueden ser distinguidas las funciones de fusión multimodal (que afectan a la interpretación) y fisión multimodal (asociadas a la generación) [122]. La fusión multimodal consiste en la combinación en un flujo de datos único de las distintas expresiones que ofrece el usuario a través de las diferentes modalidades y lenguas. Por su parte, la fisión multimodal se encarga de realizar la función complementaria sobre las contribuciones del sistema, transformando su flujo de datos de salida único en expresiones completas y coordinadas que se producen a través de las distintas modalidades y lenguas disponibles.

Tanto la fisión como la fusión multimodal surgen de la incorporación de habilidades multimodales y multilingües a la interacción. Estas son las funciones que hacen transparente al núcleo del sistema de interacción natural aquellas modalidades y lenguas en que los usuarios desarrollan la interacción, permitiendo independizar la solución obtenida de idiomas particulares o de canales concretos de expresión.

Fusión Multimodal

La fusión multimodal se encarga de recoger las salidas producidas en paralelo por los distintos componentes de la interfaz de entrada y de integrarlos a nivel léxico o semántico en un único flujo de datos. El resultado de la fusión multimodal es el equivalente a una interacción monolingüe desarrollada a través de una única modalidad. Esta representación resulta más fácilmente procesable por el sistema.

La fusión multimodal puede ser realizada a nivel de señal (*fusión léxica*); a nivel semántico (*fusión semántica*); o siguiendo una aproximación intermedia (*fusión híbrida*).

La *fusión léxica* parte de datos de cada una de las modalidades de entrada sin interpretar semánticamente y las combina en un vector único al que un clasificador, previamente entrenado,

permite dotar de contenido semántico. El resultado es una secuencia de datos expresada en una unidad de reconocimiento única (por ejemplo, fonemas). Para la implementación de este tipo de clasificadores es frecuente aplicar Redes Bayesianas, como en el caso de RoboX [94], o Redes Neuronales, que pueden ser entrenadas por separado para determinadas modalidades, como las producidas por el reconocimiento acústico y visual [87]. Las principales desventajas de este enfoque son su elevado coste computacional (para la fase de entrenamiento) y que su planteamiento es, en la práctica, poco escalable.

Por su parte, la *fusión semántica* se realiza una vez han sido interpretados por separado los fragmentos de información que provienen de diferentes modalidades (por ejemplo, audio y vídeo). Sólo tras haber sido interpretado cada uno de los flujos de datos por separado y haber sido descompuestos en sus unidades propias (fonemas para el audio, visemas para el vídeo, etc.), son integrados en una secuencia de datos única de representación uniforme (por ejemplo, en forma de fonemas). Así, la fusión se realiza tras haber sido aisladas las características relativas a cada modalidad del resto de las señales físicas originales, por lo que se hace posible que los distintos reconocedores puedan ser entrenados por separado y se simplifique el desarrollo del componente de fusión. El resultado es una aproximación más sencilla y escalable que la anterior. Un ejemplo concreto de este tipo de fusión es el descrito por Liu et al. [119], en el que se utilizan Modelos Ocultos de Markov Acoplados para la fusión de datos acústicos y visuales. El ejemplo se apoya en modelos de estados asíncronos y permite conservar las dependencias audiovisuales naturales.

Existen dos formas diferentes de aplicar fusión semántica sobre los datos de entrada:

- *En paralelo*: Consiste en realizar la integración de los flujos de datos con una granularidad temporal pequeña.
- *En serie*: Se aplica una granularidad temporal mayor, de forma que los flujos de información se concatenan secuencialmente.

Cuando el sistema no cuenta con un componente específico para la interpretación de diálogo existe una forma adicional de aplicar la fusión semántica: *la contextual*. En este caso no se consideran restricciones temporales, sino que los flujos de datos se combinan directamente con los resultados acumulados en el contexto (representado en el Modelo de Sesión [Apartado 4.2.3.2]). Se analiza si los nuevos resultados son complementarios con alguno de los resultados ya incluidos en él y, de ser así, se incorporan al contexto.

La tercera estrategia de fusión, la denominada *fusión hibrida*, combina los enfoques léxicos y semánticos. Este tipo de fusión suele aplicarse para la fusión de datos acústicos y

visuales [148]. Consiste en realizar una clasificación léxica previa de la combinación de señales procedentes de diversas modalidades (por ejemplo, audio y vídeo), obteniéndose un flujo de datos único (fonemas) que será integrado al flujo de observaciones de otra modalidad (por ejemplo vídeo, previamente representando en forma de visemas) y simplificar, por último, ambos flujos de datos en un único flujo de datos de salida (expresado, por ejemplo, a partir de fonemas). Este tipo de fusión permite obtener mejores prestaciones que las aproximaciones léxicas o semánticas por separado.

Aunque el componente de fusión multimodal no está entre los principales responsables de un desarrollo temporal avanzado de la interacción, debe concebirse de forma que no suponga un obstáculo para la interpretación incremental de las contribuciones producidas por los usuarios. Dicho de otro modo, la fusión multimodal debe realizarse en tiempo real, generando resultados parciales a medida que se desarrolla la enunciación del usuario, y no sólo tras haber sido completada. Con este fin, serán preferibles fusiones semánticas o híbridas realizadas con granularidad temporal baja. De otro modo no resultaría posible reproducir muchos de los fenómenos de la *interacción natural*, como son la producción de realimentación simultánea, o la correcta gestión de las interrupciones y los solapamientos.

Fisión Multimodal

El componente de fisión multimodal recibe el flujo de datos correspondiente a las contribuciones generadas por el sistema y lo transforma en respuestas multimodales. Para cada posible modalidad, genera un conjunto de expresiones de tal forma que el conjunto tiene significado equivalente al flujo de datos recibido del núcleo del sistema. Del mismo modo, este componente es el responsable del diseño de la presentación de la expresión del sistema, definiendo como se coordinan las distintas expresiones entre ellas y en qué momento serán producidas.

Las contribuciones del sistema deben ser adecuadas a las circunstancias en las que se producen. Esto requiere considerar el estado emocional de los participantes (nivel de formalidad de la conversación, intimidad, etc.), el conocimiento de la situación (si está permitido hablar en el lugar donde se desarrolla la interacción, el idioma que debe utilizarse, etc.), las características específicas de la sesión (las modalidades que participan en la interacción, el tamaño de la pantalla, etc.) y qué usuarios participan en la interacción (quienes tendrán sus propias preferencias). Para ello, el Adaptador Multimodal aplica los modelos que representan en el sistema estos tipos de conocimiento [Apartado 4.2.3] en combinación con las propias heurísticas que él mismo ha desarrollado (en base su experiencia sobre el éxito o fracaso de la aplicación de distintos recursos expresivos en distintas circunstancias). Un ejemplo es el generador de

respuestas del sistema Olga [11], que utiliza un conjunto de reglas para elegir la modalidad de expresión de sus contribuciones.

Para abordar una toma de turno avanzada será preciso prestar especial atención a determinados casos de expresión facial y movimiento de manos y cabeza que permiten regularla (solicitud, cesión y mantenimiento de la palabra).

4.2 *GESTOR DE INTERACCIÓN*

El Gestor de Interacción es el componente responsable del comportamiento interactivo del sistema. Su núcleo es el Gestor de Diálogo, componente que representa el progreso de las *metas compartidas* por los participantes, así como de sus propias metas discursivas (*metas individuales*). En este trabajo se aplica un modelo específico de diálogo: El Modelo de Hilos. Se trata de un *modelo intencional de acción combinada* [Apartado 3.1.4] que articula la interacción en forma de *hilos de diálogo*.

El Gestor de Interacción también incluye el componente encargado de aplicar durante la interacción la tecnología a la que los usuarios desean acceder (habilidades no interactivas del sistema, como aplicaciones, manejadores, etc.), y el conjunto de componentes que gestionan todo aquel conocimiento sociolingüístico que puede ser utilidad en la interacción. Estos son, respectivamente, el Modelo de Tareas, para representar el conocimiento operativo en los Sistemas de Interacción Natural, y los Modelos de Usuario, Sesión, Situación, Emocional y Auto Modelo. Tanto estos componentes, como el Gestor de Diálogo, serán descritos a lo largo del presente apartado.

4.2.1 Gestor de Diálogo

El Gestor de Diálogo es el componente responsable del comportamiento interactivo del sistema. En este sentido, se encarga de organizar la interacción a nivel global y local (relación entre las metas de la interacción y el desarrollo particular de cada una de ellas) [Apartado 2.3]. Para ello, desempeña funciones asociadas a la representación del dominio y del estado de la interacción, y a la interpretación y generación de diálogo:

- *Representación del Dominio y Estado de Interacción*: Modela los diálogos que el sistema es capaz de desarrollar y el estado en el que se encuentran los diálogos concretos que está manteniendo el sistema en un determinado instante.

- *Interpretación de Diálogo*: Se encarga de actualizar el estado de interacción y el conocimiento sociolingüístico asociado (principalmente el contexto de la interacción) a partir de los flujos de datos de entrada recibidos del usuario.
- *Generación de Diálogo*: Decide qué movimientos debe producir el sistema en la interacción y determina cuáles son las enunciaciones a través de las cuales se desarrollarán estos movimientos.

Para desarrollar tales funciones, el Gestor de Diálogo aplica el conocimiento recogido en los diferentes modelos de conocimiento sociolingüístico (sesión, situación, etc.) [Apartado 4.2.3], y se vale de los servicios ofrecidos por el Gestor de Tareas [Apartado 4.2.2] para llevar a cabo la resolución de los problemas que se plantean en los distintos estados de interacción alcanzados.

4.2.1.1 Representación del Dominio y del Estado de Interacción

El modelado del diálogo implica representar, tanto el conocimiento necesario para dialogar dentro de un dominio de interacción concreto, como el conocimiento sobre qué intenciones se desarrollan en un momento concreto de la interacción y el estado en el que se encuentran.

Dadas las necesidades de la propuesta de tratar la interacción como una acción combinada (por depender de ello el poder dotar al sistema de la capacidad para gestionar el desarrollo temporal de la interacción de forma realista), se ha aplicado el modelo de diálogo intencional de acción combinada denominado Modelo de Hilos [21], que se basa en la noción de *hilos de diálogo*.

Hilos de Diálogo

Durante la interacción los participantes desarrollan hilos. Se parte de la idea de que la interacción consiste en el desarrollo de las intenciones que los participantes introducen en la interacción para satisfacer sus metas. Según este modelo, un hilo representa los posibles desarrollos de una intención concreta. Es decir, la organización local de la interacción.

La apertura de un hilo ocurre cuando uno de los participantes introduce la intención ésta que representa para satisfacer alguna meta propia. Una vez abierto, el hilo se desarrolla hasta que se alcanza su resolución (exitosa o no). El desarrollo de un hilo consiste en su transición de unos estados a otros. En ellos podría ser necesaria la resolución de tareas. Las transiciones de unos estados a otros son desencadenadas por las contribuciones que realizan los distintos participantes en la interacción, o como el resultado de alguna tarea (evento). De esta forma, se

definen las distintas secuencias de juegos de diálogo que pueden hacer progresar una intención. Por ejemplo, cuando alguien tiene la necesidad de conocer la hora (meta) podría introducir en el diálogo la intención de conocer la hora "*¿Tienes hora?*", a lo que su interlocutor podría responder "*son las tres y cuarto*" o "*no llevo reloj*", describiendo estos los posibles juegos de diálogo que pueden darse en la resolución de la intención (y, en conjunto, definiendo el hilo que permite resolver dicha meta).

De los posibles resultados del desarrollo de un hilo pueden alcanzarse finales preferidos y no preferidos, en función de que permitan resolver de forma exitosa o no la meta que los motiva. De esta forma "*son las tres y cuarto*" permitiría alcanzar un final preferido al desarrollo de la meta de conocer la hora, mientras que "*no llevo reloj*" alcanzaría un final no preferido.

Los Hilos como Acción Combinada

Dado que las acciones producidas en el marco de una interacción son en sí mismas acciones combinadas (resultantes de la combinación de las acciones individuales de los distintos participantes), los hilos constituyen también acciones combinadas, por lo que tienen distintas caras. Estas caras son las acciones individuales de cada uno de los participantes que las comprometen (*hilos individuales*) y la acción combinada de todas ellas (*hilo combinado*). Inicialmente un participante introducirá un hilo (individual) en la interacción para satisfacer una meta propia, si el resto de participantes acepan el desarrollo de dicha intención, dicho hilo individual será complementado con los hilos individuales de los distintos participantes sobre la misma intención, alcanzándose un hilo combinado. Los participantes acceden a desarrollar las metas que proponen otros en la medida en que les permite también alcanzar las suyas propias.

Por ello, cada participante deberá mantener una representación de sus hilos individuales (aquellos introducidos para satisfacer sus propios intereses), de sus conjeturas sobre los hilos individuales del resto de participantes (los introducidos por otros participantes), y de sus conjeturas sobre los hilos combinados que se desarrollan en la interacción (el compromiso alcanzado entre hilos individuales de distintos participantes). Esto es lo que se denomina *zona común* [Apartado 2.2.1].

Se parte de la suposición de que tanto el usuario como el sistema son capaces de introducir nuevos hilos (*interacción de iniciativa mixta*). En el caso del sistema, estos serán introducidos cuando:

1. Cualquier componente de la interacción precise, para el mejor desempeño de sus funciones, del desarrollo de un fragmento de interacción (subdiálogo) con el interlocutor.

2. Algún componente externo decreta esa necesidad, como resultado de un proceso o por el reconocimiento de un evento. Por esta última causa se puede iniciar incluso un diálogo completo.

Gestión del Compromiso

Al tratarse un hilo de una acción combinada, está caracterizado por una medida del *compromiso* asumido por los participantes a desarrollarlo. Este compromiso es establecido en torno a la atención que los participantes prestan al hilo; el interés que tienen en desarrollarlo y la calidad del conocimiento mutuo compartido sobre é. En la medida en la que un interlocutor perciba una buena sintonía entre todos los participantes y en la que éstos se muestren favorables a desarrollar una intención, el Gestor de Diálogo asignará a su representación del hilo combinado valores elevados de compromiso en cada una de estas facetas. Sin embargo, si alguno de los participantes denota pérdida de interés o de atención sobre el desarrollo de un hilo (o si entiende que la información compartida sobre el hilo es insuficiente o de poca calidad), dichos parámetros se resentirán. De esta forma, a medida que los participantes contribuyen (o no) a desarrollar un hilo, su nivel de compromiso evoluciona, permitiendo mantener una medida de su salud.

Los participantes tienen intereses en el desarrollo de las intenciones que han comprometido. Por ello, cuando la salud de un hilo se degrada, es necesario desarrollar estrategias orientadas a recuperarla. En estos casos serán aplicados *hilos de refuerzo y reparación* para recuperar la salud el hilo degradado. Cada uno de estos hilos permitirá desarrollar técnicas específicas encaminadas a mejorar la salud del hilo. Entre estas técnicas se encuentran las reafirmaciones gestuales, los anuncios, confirmaciones implícitas y explicitas, o incluso las preguntas directas.

En función del nivel de degradación de un hilo y de la naturaleza de dicha degradación, podrán ser aplicadas unas técnicas u otras, cada una de ellas con un impacto distinto sobre la interacción. Gracias a la aplicación de estas técnicas es posible hacer ganar atención sobre el hilo frente al resto de hilos desarrollados, mejorar el interés de los participantes en desarrollarlo, y solucionar confusiones en la información compartida.

La Estructura Intencional

Al igual que el concepto de hilo determina la organización local de la interacción, la Estructura Intencional representa la organización global: Los hilos mantienen unas relaciones de dependencia entre ellos que requieren ser también representadas en el estado de interacción.

Según el Modelo de Hilos, la organización intencional se define como una estructura arbórea donde cada hilo es hijo de otro hilo, y a la vez es padre de aquellos hilos que desarrollan subintenciones de su intención principal. De esta forma, las intenciones de la interacción tiene una organización jerárquica, y cada una de ellas será parte de una intención superior, que responde a propósitos más generales. La intención más general que se puede dar es completar la conversación en sí misma, que se representa por el *hilo base* y constituye la raíz de toda interacción. Cualquier otro hilo introducido será descendiente, directo o indirecto, del hilo base.

Cada vez que un nuevo hilo individual aparece, es ajustado a la estructura arbórea de hilos combinados que mantiene el sistema. Este ajuste consiste en identificar de qué intención de la estructura intencional depende la intención del nuevo hilo, o si se trata de una intención que ya está siendo desarrollada en la interacción. El ajuste puede conllevar el refuerzo del hilo actual, la reapertura de un hilo que se consideraba cerrado, la fusión de un hilo anterior con uno nuevo, la inicialización de un nuevo hilo, el cierre de un hilo existente, etc.

Este tipo de representación de la organización global tiene como principal ventaja el permitir mantener abiertas distintas líneas de discurso durante la conversación (cada una de las cuales es un hilo), que podrán ser atendidas en ordenes diversos, según requiera la interacción en cada momento.

La Estructura Focal

El orden en que se resuelven las distintas metas introducidas a lo largo del diálogo no es aleatorio, igual que ocurre en las conversaciones entre personas. En todo momento, la atención se centrará sobre alguno de los hilos que se encuentran abiertos y ambos interlocutores centrarán su atención en él con un determinado nivel de compromiso.

En la interacción los hilos pueden ser desarrollados, cancelados, o abandonados temporalmente y reabiertos más tarde, siendo el *hilo enfocado* aquel sobre el que se centra la atención en un determinado momento de la conversación. Sobre él se exige coherencia intencional, contextual y estructural.

Cuando una nueva intención aparece en el diálogo, el hilo enfocado que se viene desarrollando puede ser cuestionado y desplazado para enfocar esta nueva intención. El hilo que se estaba desarrollando hasta el momento queda pendiente de continuar su desarrollo y se pasa a desarrollar el nuevo. Cuando un hilo llega a su estado final se resuelve, y los participantes de la interacción volverán a buscar entre todos los hilos abiertos un nuevo foco. El participante que ostente el turno escogerá la opción más plausible (o la que más convenga individualmente) de entre los hilos pendientes de desarrollar. De hecho, cualquier participante puede cambiar el hilo

enfocado durante el ejercicio de su turno según sus intereses, incluso si el hilo anteriormente enfocado no hubiera sido resuelto. En definitiva, al cambiar de foco, se pasaría a seguir desarrollando una meta que quedo pendiente. Incluso se podrán reabrir hilos ya cerrados para refinar resultados.

A lo largo de todo el proceso comunicativo, el receptor debe esforzarse en seguir el plano de interpretación de su interlocutor. Por su parte, el emisor debe contribuir, en la medida de lo posible, a que al receptor le resulte suficientemente fácil seguir la conversación (siguiendo ciertos convenios en la atención, como el orden de introducción o el compromiso mutuo, y anunciando convenientemente los cambios en la atención a otras metas). El objetivo es que ambos alcancen el mismo plano de interpretación. En la medida en que alcancen una sintonía en la atención sobre los hilos, la salud de la interacción se mantendrá fuerte.

El Papel del Modelo de Hilos en una Toma de Turnos Avanzada

La aplicación del Modelo de Hilos posibilita una gestión flexible y versátil del diálogo, al considerar la interacción como una acción colaborativa entre los participantes y no como una secuencia predefinida de pasos a seguir (*gramáticas de diálogo*), el rellenado de formularios (*marcos*) o el obligado desarrollo de las intenciones introducidas unilateralmente por alguno de los participantes (*modelos intencionales*). Todo esto, junto a la gestión del compromiso alcanzado sobre los hilos, ofrece numerosas ventajas en lo que a la naturalidad de la interacción se refiere frente al resto de modelos.

Más concretamente, esta forma de gestión del diálogo es buena para una toma de turnos avanzada porque permite introducir y desarrollar en cualquier momento nuevos hilos, permitiendo la ocurrencia de interrupciones y auto interrupciones para reenfocar hilos distintos al enfocado, o para insertar otros nuevos (sin que por ello sean descartados los progresos alcanzados en el desarrollo de los hilos interrumpidos). Además, aporta herramientas para una gestión adecuada de la atención, intención e interés (por representar las dependencias intencionales entre los hilos a través del estado intencional, el orden de prioridad en la atención en la estructura focal y el compromiso alcanzado en el desarrollo de cada hilo), lo que facilitar la detección y reparación de problemas de sintonía entre los participantes.

4.2.1.2 Interpretación de Diálogo

La función de interpretación de diálogo consiste en analizar las implicaciones que tienen en la interacción cada una de las distintas enunciaciones emitidas por el resto de participantes a los niveles estático, dinámico y estructural. La interpretación de una enunciación conllevará la

actualización del estado de interacción y del conocimiento manejado por el sistema en el contexto.

De forma genérica, la interpretación se realiza a partir del flujo de datos que recibe el Gestor de Diálogo y que representa una enunciación de usuario expresada en forma de secuencias de actos comunicativos. Durante la interpretación, la secuencia de actos comunicativos deberá ser sometida a los análisis que permiten resolver las referencias deícticas que ésta pudieran contener. También deberán ser identificados los movimientos que deben producirse en el diálogo como consecuencia de tales actos comunicativos (*interpretación estructural*), deberá resolverse si conllevan la inserción de nuevas metas (*interpretación dinámica*), y deberá extraerse la nueva información que aporte cada uno de los actos comunicativos para actualizar la información sociolingüística (*interpretación estática*).

4.2.1.3 Generación de Diálogo

Por su parte, la generación de diálogo consiste en la producción de los comportamientos del sistema que tienen sentido dentro del estado de interacción, aplicando el conocimiento que proviene de los componentes del contexto y las tareas realizadas por el gestor de tareas. La generación conllevará el desarrollo de nuevos movimientos en la interacción (*generación estructural*), la inserción de nuevas iniciativas (*generación dinámica*) y también la actualización del contexto (*generación estática y emocional*). El resultado será expresado en forma de una secuencia de actos comunicativos que, tras la aplicación de los procesos pertinentes de generación de referencias deícticas en caso de ser necesario, el gestor de diálogo pondrá a disposición del Gestor de Presentación para desencadenar la enunciación del sistema.

4.2.2 Gestor de Tareas

Los sistemas de diálogo son el interfaz que conecta a los usuarios con la tecnología a través de una interacción más natural. El objetivo es facilitar a las personas la resolución de sus problemas en los escenarios específicos en los que podrá desarrollarse la interacción. Desde el punto de vista del Gestor de Interacción, estos problemas serán resueltos como tareas por otras aplicaciones, generalmente externas al sistema, que funcionan a modo de proveedores de servicio. De esta forma, la gestión del diálogo debe complementarse con un elemento a través del cual puedan ser solicitados dichos servicios cuando el estado de interacción lo requiera. Este elemento será el Gestor de Tareas.

Por medio del Gestor de Tareas, los Sistemas de Interacción Natural desarrollarán las habilidades específicas que requieren los usuarios para la resolución de sus tareas a través de

componentes externos. Estas habilidades varían en función del dominio de interacción, por lo que es de gran importancia que la arquitectura de los gestores de tareas sea fácilmente adaptable a aplicaciones externas distintas. Es decir, los gestores de tarea deben permitir incorporar fácilmente nuevas habilidades cuando así lo requiera el dominio de aplicación. Por ello deben ser descompuestos en dos partes bien diferenciadas: un *motor de razonamiento* (que aplica los algoritmos lógicos y las condiciones a la resolución de las distintas tareas que se pretenden resolver); y una *capa de persistencia* (que da acceso a las aplicaciones y permite controlarlas).

Respecto a la capa de persistencia, la resolución de las tareas como solicitudes de servicio a otros componentes externos pasa por describir: la organización de las tareas y las relaciones que existen entre ellas [138]; establecer la correspondencia entre la ontología interna del sistema y la de cada una de las aplicaciones externas que podrán ofrecer servicios; y determinar las distintas formas en las que se pueden invocar estos servicios y los posibles resultados que pueden devolver.

En general, la adaptación de aplicaciones externas debe hacerse de forma individualizada para cada una de las posibles tareas a ofrecer en el dominio de interacción. Sin embargo, cada vez más los fabricantes implementan estándares de comunicación que permiten el control de sus productos a través de interfaces de ámbito general. Tal es el caso de la familia de estándares Universal Remote Console, URC [176], que permite incluir interfaces estandarizados de acceso a los productos a través de los cuales pueden ser utilizados con dispositivos como los teléfonos inteligentes o los Sistemas de Interacción Natural. De esta forma, se abre la posibilidad de implementar capas genéricas de acceso a las tareas, a través de las cuales los dispositivos puedan acceder a los servicios presentes en el entorno (por ejemplo, cafeteras, alarmas, vídeos, termostatos, electrodomésticos, etc.) sin necesidad de interfaces específicos para cada uno de ellos. No obstante, quedan todavía algunos problemas pendientes, especialmente los relacionados con la generalización ontológica.

Los componentes internos del sistema también pueden ofrecer servicios a la gestión del diálogo a través componente gestor de tareas. Por ejemplo, los distintos modelos de conocimiento sociolingüístico relacionado con la interacción aportan información del usuario, sesión, situación, personalidad del sistema y estado anímico de los participantes al desarrollo de la interacción. Algunos modelos, como el Modelo de Situación [Apartado 4.2.3.3], pueden ofrecer otro tipo de servicios más complejos, como la descripción de los pasos necesarios para pasar de una situación a otra o la descripción de una determinada situación (servicios ambos que pueden ser utilizados para ayudar al usuario cuando requiera tal información, o pueden ayudar al sistema a planificar las estrategias de diálogo que seguirá). Entre las tareas internas al

componente de interacción, pueden encontrarse también efectos que supongan modificaciones sobre el contexto, el estado de interacción, su curso e incluso el diseño del diálogo. Este tipo de servicios serán muy dependientes de la aproximación en que se base el componente Gestor de Diálogo.

4.2.3 Conocimiento Sociolingüístico Asociado a la Interacción

Junto al Gestor de Diálogo y el Gestor de Tareas [Apartados 4.2.1 y 4.2.2], existe todo un conjunto de componentes adicionales que representan el resto de conocimiento sociolingüístico que entra en juego en el desarrollo de la Interacción Natural. Entre ellos se encuentra el conocimiento relacionado con el usuario, la información de lo ocurrido hasta el momento en la interacción, la situación, los afectos puestos en juego y la propia personalidad del sistema. Respectivamente, cada uno de estos aspectos del conocimiento es gestionado por los siguientes componentes: Modelo de Usuario, Modelo de Sesión, Modelo de Situación, Modelo Emocional y Auto Modelo.

4.2.3.1 Modelo de Usuario

El Modelo de Usuario es el componente que representa la información relativa a los posibles usuarios del sistema. Aunque la información de los usuarios puede ser alterada durante la interacción, su validez se extiende más allá de la propia conversación. El objetivo principal de este modelo es personalizar la interacción para los usuarios concretos (o grupos de usuarios) y ajustarla, en la medida de lo posible, a sus preferencias, limitaciones y experiencia. Es por esto que, a medida que se han incrementado los esfuerzos en la mejora de la adaptabilidad de los Sistemas de Interacción Natural, este es uno de los componentes que ha ganado un mayor interés.

Existen dos posibles enfoques para los modelos de usuario. El primero modela las características de usuarios concretos [95] y el segundo utiliza modelos de grupos [18]. El modelado de las características de usuarios individuales consiste en la representación de información descriptiva sobre el propio usuario y sobre la relación que existe entre este y los elementos de la aplicación. Con ello se pretende conseguir la adaptación en términos de presentación, técnicas comunicativas utilizadas, contenidos presentados y navegación. Este enfoque, al almacenar información específica para cada individuo, resulta ser poco respetuoso con la privacidad de los individuos y difícilmente escalable. El modelado de grupos de usuarios, por su parte, es un enfoque respetuoso con la privacidad de los usuarios y permite reducir la cantidad de datos almacenados cuando el número de usuarios crece (en comparación con el enfoque a características de usuarios individuales). Según Lenzmann y Wachsmuth [114] el

objetivo no es tanto averiguar cuál es el usuario, sino saber cuáles son su características para poder desarrollar una interacción que se adapte a él. En este caso, más que disponer de un perfil específico para cada posible usuario, el sistema modela perfiles de arquetipos de usuarios. De esta forma, el principal problema es cómo identificar durante la interacción el grupo específico al que pertenece un determinado usuario. Para resolver este problema se aplican técnicas de clasificación (k vecinos cercanos, árboles de decisión, etc.) sobre las observaciones del usuario durante la interacción. A la labor contribuyen los códigos y símbolos que utiliza, su personalidad (estados de ánimo y su evolución), características dialógicas (tendencia a interrumpir, realimentar, etc.), uso de indirecciones e ironías, problemas que pretende resolver a través del diálogo, etc.

El grupo al que se asocia un usuario no tiene por qué ser el mismo durante toda la interacción, y de hecho varía dinámicamente a medida que la interacción aporta nuevos datos sobre el usuario. Los grupos de usuarios están estructurados de forma jerárquica, y van desde los grupos más generales (que no aportan demasiada información sobre los usuarios) hasta los más específicos (donde el conjunto de características que permiten asociar a un usuario es más elevado y específico). El usuario comenzará perteneciendo a un grupo muy genérico y, en función de la información que vaya apareciendo sobre él en la interacción, podrá concretarse dinámicamente el subgrupo al que pertenece. La cantidad de información de la que se dispone de un determinado usuario incrementa a medida que puede ser encajado en grupos más específicos.

Dado que la elección del grupo al que pertenece un usuario es una estimación, se hace necesaria una medida de certeza a la validez de las características supuestas para un determinado usuario. Cuando esta certeza es demasiado baja y se trata de características importantes para la interacción en un determinado momento, será necesario aplicar técnicas de refuerzo y reparación para mejorar la calidad de información disponible sobre el usuario. Estas técnicas consisten en interrupciones, confirmación explícita o implícita, realizadas por medio de la inserción de eventos en el Gestor de Diálogo [Apartado 4.2.1].

4.2.3.2 Modelo de Sesión

El Modelo de Sesión es el componente que gestiona el conocimiento relativo al desarrollo de la interacción. Esta información tiene una validez limitada a la duración de la interacción y puede estructurarse en información estática de la sesión; historia de la sesión y el conocimiento para la resolución de referencias deícticas y anafóricas.

Información Estática de la Sesión

La información estática de la sesión está compuesta por partículas atómicas que recogen los datos que han sido aportados por los participantes a lo largo de la interacción. Estas partículas atómicas se denominan líneas de contexto y son almacenadas con vínculos a las enunciaciones en la que se aportaron. Cada línea de contexto se refiere a uno de los conceptos recogidos en la ontología del sistema de interacción, asociándole un valor específico y con unas determinadas garantías de confianza en la fuente (probabilidad de que el interlocutor que aporta la información la conozca realmente) y veracidad (medida subjetiva de la credibilidad en la intención del interlocutor en aportar una información verdadera).

Durante la interacción pueden registrarse líneas de contexto referentes al mismo concepto, pero con distintos valores. La causa puede ser la modificación de dicho valor durante la interacción (en este caso conocer la antigüedad de las líneas de contexto será de gran utilidad), que se trate de un concepto que admita múltiples valores ("*mi helado es de fresa y nada*"), o por la ocurrencia de contradicciones en la enunciación ("*creo que lloverá o nevará*"). La credibilidad en la fuente y la veracidad de la información serán imprescindibles para resolver este tipo de conflictos.

Dado que la conversación admite una descomposición en diálogos y subdiálogos, la organización de la información estática de la sesión es, a su vez, jerárquica. Así, en ocasiones la información estática relativa a cada subdiálogo puede también formar parte del diálogo al que pertenece (*herencia abajo-arriba*) y, salvo colisión, la información estática de un diálogo es aplicable a sus subdiálogos (*herencia arriba-abajo*). Los procesos de herencia reducen la credibilidad de las líneas de contexto heredadas, y este efecto es mayor en los procesos de herencia arriba-abajo.

La recuperación de la información estática relativa a un concepto puede realizarse sobre la totalidad de las líneas de contexto que se refieren a dicho concepto, con respecto a la línea de contexto de mayor credibilidad, o imponiendo un valor mínimo de credibilidad. En este último caso, cuando no existe ninguna línea de contexto de credibilidad suficiente, se introducirán sobre el gestor de diálogo eventos encaminados a reforzar la credibilidad de la información. Las técnicas más comunes aplicadas en la interacción humana son el anuncio (expresar de forma explícita la información que se va a utilizar para dar la oportunidad a los interlocutores de corregirla en caso de no ser correcta) y la introducción en la interacción de subdiálogos de solicitud de confirmación de la certeza de dicha información.

Historia de la Sesión

Es la parte del Modelo de Sesión que recoge las enunciaciones acaecidas en el transcurso de la interacción. Pueden ser almacenadas tal y como se produjeron, de forma literal, o tras haber sido procesadas desde un punto de vista semántico. La historia de la sesión resulta de gran utilidad para:

- Llevar a cabo reparaciones en los casos de haberse realizado una interpretación incorrecta de alguna enunciación de usuario. Recuperándola, el error podrá ser corregido en los casos en los que posteriormente se haya conseguido nueva información relativa a dicha enunciación o solicitando al usuario que la reformule en caso contrario.
- Conseguir mayor variabilidad en la interacción, evitando comportamientos de apariencia mecánica.
- Resolver anáforas, ya que consisten en referencias a elementos mencionados anteriormente en la interacción.
- Aplicar estrategias dialógicas alternativas cuando el sistema se encuentre en situaciones confusas. La aplicación de estrategias alternativas sobre el conjunto de las enunciaciones ya desarrolladas puede ofrecer un mejor resultado.

Resolución de Referencias

En ocasiones, la resolución de los conceptos a los que hacen referencia algunos de los símbolos que aparecen a lo largo de la interacción, no puede realizarse sin tener en cuenta la situación en la que ésta se desarrolla y el resto de la información que ha sido aportada durante la interacción. Se trata de casos en los que los símbolos apuntan a alguna entidad lingüística anterior ("*Mercedes se lo dijo a Pedro*"), posterior ("*A esto me refiero: te has portado mal*"), o a entidades no lingüísticas ("*está aquí*", "*lo vi ayer*", etc.). Estas son, respectivamente anáforas, catáforas y deixis.

Las anáforas y catáforas consisten en la sustitución de expresiones correferenciales próximas por términos deícticos. La anáfora puede afectar a la enunciación actual, en cuyo caso puede resolverse durante la interpretación de lenguaje natural, o pueden involucrar enunciaciones anteriores o posteriores (fenómenos referenciales cruzados) del mismo participante o de otros, y por tanto requerir la consulta de la historia de la sesión. En el caso de la catáfora, si la referencia es sobre la enunciación actual podrá ser fácilmente resuelta durante la interpretación de la intervención, aunque si entran en juego enunciaciones futuras (caso menos frecuente), será más difícil de detectar y resolver. También pueden ocurrir anáforas y catáforas que vinculen expresiones de distintas modalidades (anáforas y deixis cruzadas).

El resto de fenómenos deícticos pueden consistir en referencias a personas, momentos, discursos, objetos del espacio, y también pueden ser sociales:

- La deixis de persona puede ser resuelta en la fase de procesamiento de lenguaje natural.
- Para la resolución de la deixis a momentos, la deixis temporal, debe ser tenido en cuenta tanto el momento en el que transcurre la interacción, como cuándo ocurrió la propia enunciación deíctica y la capa (de las líneas de acción) a la que pertenece dicha contribución. Esta información es la que permite fijar la base de tiempos en la que la referencia tomará sentido.
- La deixis de discurso ("*como dije antes*") se trata de referencias a porciones de discurso que se usan para retomar el desarrollo de un segmento del discurso y heredar parte de su información estática.
- La deixis a objetos del espacio son referencias que sólo toman sentido en el lugar en el que se desarrolla la interacción, por lo que involucrarán al componente de situación. Este tipo de referencias suelen ser realizadas a través del lenguaje, aunque muy frecuentemente son acompañadas por gestos y miradas.
- La deixis social agrupa a aquellos elementos que describen las relaciones sociales entre los interlocutores que determinan aspectos tales como el grado de respeto o intimidad, distanciamiento o insulto, etc. Para su tratamiento se requiere también la contribución del modelo de situación.

La ocurrencia de referencias en la interacción afecta tanto a la interpretación de las enunciaciones de otros participantes como a la generación de las contribuciones propias del sistema. Por economía del lenguaje y con el fin de alcanzar una interacción más natural, el sistema debe también valerse de estos fenómenos en sus enunciaciones.

4.2.3.3 Modelo de Situación

Mientras que el Modelo de Sesión [Apartado 4.2.3.2] gestiona la información interna a la propia interacción, el Modelo de Situación trata con el contexto en el que ésta se desarrolla. Es el modelo que enmarca la sesión en las circunstancias que la rodean.

Los aspectos circunstanciales que deben ser considerados es un tema no libre de polémica, aunque la clasificación más extendida es la propuesta por Gee [66], que define:

- *Aspecto semiótico*: El lenguaje o los signos usados en la interacción.

- *Aspecto político*: Los roles asignados a cada uno de los interlocutores.
- *Aspecto operativo*: Operaciones permitidas dentro de la interacción.
- *Aspecto material*: Situación espacio-temporal en la que se desarrolla la interacción.
- *Aspecto socio-cultural*: El entorno social de la interacción.

El modelo de situación, al monitorizar la situación en la que se desenvuelve la interacción, puede desencadenar eventos que alteren el curso del diálogo (lanzados directamente sobre el gestor de diálogo). Por ejemplo, podrán ser iniciados subdiálogos para informar al usuario de la ocurrencia de determinados hechos ("*Ya son las ocho de la mañana*") o de que se dan situaciones propicias para realizar alguna acción ("*Acaban de abrir la cafetería, quiere que le guíe hasta allí*"). Por otro lado, el modelo de situación puede ofrecer determinados servicios de resolución de tareas, como por ejemplo aportar información de situaciones pasadas, predecir situaciones futuras, llevar a cabo planes para alcanzar situaciones de destino desde situaciones de origen, etc.

4.2.3.4 Modelo Emocional

Por emoción se entiende una "alteración del ánimo de cierta intensidad, duración y efecto sobre el individuo, que se demuestra mediante alguna conmoción somática" [21]. Trata de un cambio de estado del sujeto que afectará a sus acciones y se manifestará de forma muy marcada a través de la quinesia, el lenguaje, del paralenguaje [17; 142]. En lo referente a este último canal, se manifestará a nivel suprasegmental (en cuestiones como el tono, la energía, la temporización o el contorno), segmental (determinando la articulación y la temporización) e intrasegmental (afectando a la calidad de la voz).

La monitorización del estado emocional de los participantes puede ayudar a mejorar la interacción en numerosos aspectos. Entre otros, permite cambiar el comportamiento interactivo al observar determinados estados en el interlocutor (levantar el ánimo de un participante decaído, tranquilizar a un participante que se muestra nervioso, etc.) y ayuda a personalizar y adaptar la interacción a los participantes y su estado actual, así como las estrategias de presentación de las contribuciones del sistema. Aunque la obtención de tales resultados pasa por el tratamiento de la información emocional afectada durante la interpretación, el Modelo Emocional permite reducir la frustración producida en el usuario como consecuencia del uso del sistema, haciendo posible abordar nuevos retos interactivos.

Las emociones pueden ser clasificadas como primarias o básicas. Las primeras suelen ir acompañadas, por lo general, de manifestaciones particulares como expresiones faciales,

tendencias en el comportamiento, patrones sicológicos, etc. Derivan de necesidades evolutivas esenciales y, por tanto, son independientes de la cultura. No son exclusivas de los humanos y pueden ser experimentadas por otros mamíferos sociales. Las segundas, por su parte, consisten en un tipo de emociones más complejas que surgen por variación o combinación de las emociones primarias. Entre ellas se pueden distinguir orgullo (variación de la felicidad en respuesta a un sentimiento de éxito), gratitud (felicidad derivada de apreciar la ayuda prestada por otra persona ante una situación de ansiedad de uno) y otras como pena, afecto, sarcasmo/ironía o sorpresa/asombro.

A la hora de caracterizarlas el amplio espectro de emociones que pueden ser manifestadas por los humanos, existen dos tendencias. Una de ella consiste en identificar categorías de emociones y la otra consiste de definir las distintas dimensiones en las que se puede representar cualquier emoción. Aunque no existe un consenso en cuanto a la lista de emociones básicas que pueden ser definidas, destaca el modelo propuesto por Ekmann y Friesen [54] que distingue como emociones básicas: felicidad, tristeza, disgusto, enfado, miedo y sorpresa.

En la aproximación a dimensiones emocionales, el objetivo no es etiquetar las emociones como categorías discretas, sino caracterizar las emociones (tanto básicas como secundarias) sobre un espacio de dimensiones continuas. Las dimensiones más comúnmente utilizadas son la valentía (positiva y negativa) y el nivel de exaltación (calma o excitación), aunque según algunas investigaciones estos dos ejes podrían no ser suficientes para representar algunas emociones, como por ejemplo el miedo intenso o el enfado. En algunas ocasiones se considera la dominancia, relacionada con el sentido de control sobre el humor, como una tercera dimensión. Este tercer eje ayuda también a distinguir las emociones iniciadas por el sujeto de las obtenidas del entorno. Esta representación de las emociones tiene como principal ventaja que permite medir distancias entre categorías emocionales y transiciones graduales de una emoción a otra.

El Modelo Emocional tratará con la presentación del conocimiento del estado anímico de los participantes y de la forma en la que el estado de ánimo afectará a las enunciaciones del sistema. Las conjeturas realizadas sobre el estado emocional de los participantes deberán afectar a la forma en que el sistema se comporta en la interacción (a la interpretación de las enunciaciones del resto de participantes, a la forma en la que progresa en la interacción, y a cómo se generan y expresan las propias enunciaciones del sistema). La caracterización del estado emocional de los participantes requiere combinar el análisis de las contribuciones que realizan a través de las distintas modalidades y, también, de la forma en la que desarrollan la

interacción. El estado anímico se refleja en la prosodia de las intervenciones (tono de voz, pausas, ritmo, etc.) sobre la modalidad de habla, pero también modula el comportamiento de los usuarios a través de otras muchas modalidades como los gestos corporales, gestos faciales, movimientos de manos, dirección de la mirada, etc. Por otro lado, los movimientos que produce en el diálogo pueden ayudar a perfilar emocionalmente al interlocutor. Por último, el participante puede introducir explícitamente información sobre su estado anímico en el diálogo ("*estoy contento*").

Para desarrollar un modelado emocional más complejo, deben modelarse por separado los estados emocionales de cada uno de los participantes, incluido el del propio sistema. Deben ser definidos los posibles estados anímicos que pueden alcanzar los participantes y los rasgos emocionales de cada uno de ellos. También deben definirse las posibles transiciones de estado que pueden ocurrir y las condiciones que las desencadenan. De esta forma, se entra en el terreno de la dinámica emocional que puede desencadenar alteraciones emocionales y alteraciones en la interacción.

Las alteraciones emocionales pueden ser reacciones empáticas (buscar identificación emocional con el interlocutor), reacciones simpáticas (buscar la aprobación emocional del interlocutor), iniciativas emocionales (de origen en la personalidad del sistema), etc. Por su parte, las alteraciones en la interacción pueden ser sobre el progreso del diálogo, sobre las metas desarrolladas, etc. El estado emocional del sistema debe afectar a la forma en la que el sistema genera y expresa sus enunciaciones. En la generación, el estado anímico del sistema afectará a los patrones discursivos que aplique al construir sus enunciaciones, a la inserción de enunciaciones de fines únicamente emocionales (dirigidos a cambiar el estado emocional de los participantes) y a la forma en la que se afectan las expresiones producidas (que pueden estar restringidas en función de los patrones discursivos aplicados).

El ámbito de actuación del modelo emocional es interno a la propia interacción y, otros tipos de caracterización emocional que trasciendan sus límites serán competencia de los Modelos de Usuario [Apartado 4.2.3.1] o Auto Modelo [Apartado 4.2.3.5].

4.2.3.5 Auto Modelo

El Auto Modelo es el componente que modela la personalidad del sistema. Contiene las metas emocionales del sistema a largo plazo (por ejemplo ser cortés y amable), y también sus metas operativas (por ejemplo, promocionar los nuevos productos corporativos cuando no hay metas más importantes que desarrollar).

Desde otro punto de vista, las metas emocionales podrán ser programas o adquiridas. Aunque lo más común es dotar al sistema de unas metas emocionales estáticas de forma programada, también es posible adquirir dichas metas, de forma dinámica, en base a heurísticas sobre sus propias experiencias interactivas (por ejemplo, adaptar la relación de los pesos de las metas emocionales según mejoren o no la calidad de la interacción). Respecto a las metas operativas, del mismo modo se puede distinguir entre metas programadas y adquiridas (en cuyo caso se hablará de metas comprometidas). Serán más comunes las metas programadas, aunque el sistema también podrá comprometer con el usuario otras nuevas ("*avísame en cuanto abran la farmacia*").

El Auto Modelo, además, es el componente que almacena las experiencias personales del sistema, y que las aplica a la generación de opiniones [64].

4.3 *PLATAFORMA MULTIAGENTE ECOSISTEMA*

La Plataforma Multiagente Ecosistema [40; 175] permite modelar a cada uno de los componentes que integran la arquitectura como uno o varios agentes, en función de las necesidades específicas de cada componente. Cada agente estará capacitado para ofrecer determinados servicios a la comunidad, pudiendo existir agentes distintos capaces de servir los mismos servicios a través de estrategias alternativas [Figura 11].

Una de las características más destacables de la plataforma es su capacidad para regular por sí misma la población de agentes que contiene. Para ello está dotada de *agentes específicos para el control de la población* y de unos elementos, denominados *agencias*, a los que se adscriben los agentes de la plataforma [a, b y c]. Las agencias consisten en descripciones de agentes, donde se definen los servicios que ofrecerán al resto y el número máximo y mínimo de instancias (agentes) que estarán disponibles en cada momento. Aunque todos los agentes de una agencia comparten la misma lógica, sus decisiones no tienen por qué coincidir, puesto que pueden estar influenciadas de la experiencia individual de cada uno de ellos. A partir de la información contenida en las agencias, y de la carga que tiene cada uno de sus agentes, los agentes de control de la población crearán otros nuevos o los destruirán, en función de la carga computacional de la agencia.

En esta plataforma los clientes no solicitan servicios directamente a los servidores, ni llegan a conocer en ningún momento la identidad de los mismos. En su lugar, solicitan dichos servicios a unos agentes específicos, denominados *facilitadores* [d y e], quienes conocerán qué

agentes están más capacitados en un determinado momento para resolverlos y designarán cuáles de ellos deberán ofrecer soluciones ante la solicitud. Para la toma de dicha decisión, el agente facilitador considera parámetros diversos, entre los que se encuentra la carga computacional de cada agente.

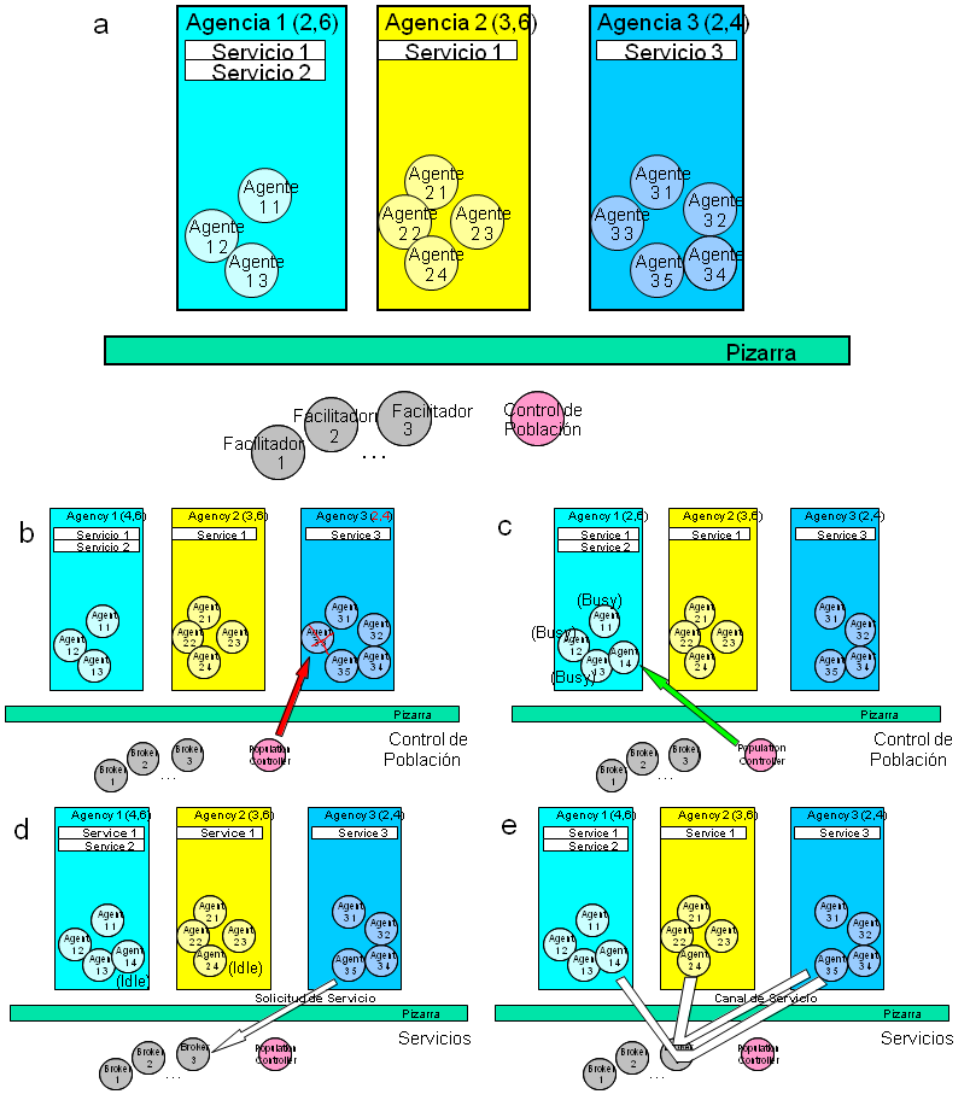


Figura 11: Plataforma Multi Agente de Pizarra Compartida Ecosistema (a) Arquitectura · (b) Eliminación del Agente 3.3 por el Agente de Control de la Población por exceso de agentes vinculados a la Agencia 3 · (c) Creación del Agente 1.4 por el Agente Control de Población por exceso de carga en los agentes asociados a la Agencia 1 · (d) Solicitud de la resolución del Servicio 1 del Agente 3.5 al Facilitador 3 · (e) Resolución del Servicio 1, solicitado por el Agente 3.5 al Facilitador 3, a través de los agentes 1.4 (de la Agencia 1) y 2.4 (de la Agencia 2), ambos disponibles en el momento de la solicitud

Los servicios llevan asociados un conjunto de parámetros que permiten caracterizar los requisitos mínimos de la solicitud. Entre ellos se encuentran la calidad mínima requerida para el

servicio (la tolerancia a imprecisiones o errores), su tiempo de caducidad (tiempo máximo que puede tardarse en obtener la solución), el tiempo de caducidad para la primera respuesta (tiempo máximo de espera para obtener una primera solución preliminar) y la criticidad de la resolución del servicio (para limitar aquellas estrategias con elevada probabilidad de no producir solución alguna).

El cliente podrá recibir, ante una misma solicitud, diferentes soluciones (cada una de ellas generada por un agente distinto aplicando estrategias diferentes). En función de la estrategia desarrollada, las soluciones podrán ser obtenidas en un tiempo mayor o menor, con distintos grados de calidad, y su obtención podrá estar o no garantizada. Del mismo modo, se acepta un refinamiento incremental de las soluciones, pudiéndose obtener desde un principio soluciones preliminares (poco precisas, pero producidas rápidamente) que se irán refinando a medida que el servidor pueda obtener resultados mejores.

Por último, la plataforma está concebida para soportar el despliegue de los agentes en redes distribuidas, permitiendo una elevada escalabilidad de la plataforma, rasgo que facilita la extensión de la implementación a medida que se la dota de nuevas capacidades interactivas.

Capítulo 5 OBJETIVOS Y PROPUESTA

Por lo general, los sistemas de interacción tienden a simplificar la toma de turno a un proceso de paso de testigo, según el cual la posesión de la palabra va pasando de un participante a otro de forma ordenada, y donde sólo el que tiene la palabra puede contribuir en la interacción y determina unilateralmente cómo lo hará (toma de turno por *ciclo de interacción* [Apartado 2.1]). Bajo este modelo de toma de turno, la validez de una contribución se mantiene desde el momento en que se formaliza hasta su completa expresión en un turno, y no requiere señalizar la intención de los participantes por ganar, mantener o liberar la atención de sus interlocutores.

Sin embargo, tal y como describen los estudios teóricos sobre el *uso del lenguaje* en la *interacción humana* [Apartado 2.2], la toma de turno no se produce como una mera consecuencia del turno previo de otro participante, y ni siquiera requiere que el canal esté libre para realizarse. La toma de turno de la *interacción natural* constituye un proceso espontáneo e impredecible, donde las contribuciones de los participantes (tanto primarias como colaterales [Apartado 2.2.3.2]) se van produciendo sobre la marcha, fruto de la confluencia de sus intereses particulares en un compromiso común [Apartado 2.2.1], y dónde los fenómenos de solapamiento e interrupción, lejos de ser anomalías, resultan ser recursos frecuentes, necesarios y naturales [Apartado 2.4]. De esta forma, los participantes, antes de comenzar su turno, sólo tienen una formalización previa de aquello que van a desarrollar, y es mientras lo van desarrollando cuando lo refinan en contenido y forma, a partir de la realimentación simultánea que reciben de los interlocutores y de los sucesivos cambios que se producen en el estado de interacción y en las circunstancias sociolingüísticas [Apartado 2.2.2]. Además, el desarrollo temporal de los turnos ofrece información de gran importancia para el correcto modelado del estado de los turnos, la posesión de la palabra y los participantes que son candidatos a tomarla. También es utilizado para aplicar estrategias de gestión de la palabra.

Los problemas que comprende una *toma de turno avanzada* pueden ser resumidos en un conjunto de cuatro funciones [173]: *Gestión de la Continuidad Temporal, Coordinación de Procesos, Gestión de la Toma de Turno* y *Gestión de Metas*. Cada una de ellas comprende los siguientes problemas:

Gestión de la Continuidad Temporal: Ni las contribuciones del sistema ni las del usuario son unidades atómicas producidas como eventos puntuales en el tiempo. Ambas tienen un desarrollo temporal y deben ser consideradas como el resultado de la combinación de diferentes fragmentos de expresiones que los participantes van produciendo a lo largo de turno. Desde el punto de vista de la interpretación, es necesario detectar la continuidad que subyace al conjunto de expresiones independientes producidas por el usuario, identificando unidades de expresión con significado interactivo completo (actos *comunicativos* [159]) susceptibles de suponer cambios en el estado de interacción. En lo referente a la generación, los cambios producidos en las circunstancias sociolingüísticas que rodean a su producción, o la interpretación de nuevas contribuciones del usuario pueden desencadenar la reformulación de las contribuciones previas del sistema, su auto interrupción, o incluso la rectificación de contribuciones previamente expresadas. Este hecho requiere la capacidad de incorporar los nuevos fragmentos de expresión producidos por el sistema a los flujos de expresión en curso de la forma más continua e hilada posible. Del mismo modo, determinados marcadores relacionados con la toma de turno son expresados como alternaciones en la continuidad temporal de la realización del turno [165]. Su detección y síntesis también forman parte de esta función.

Coordinación de Procesos: La independencia entre los distintos procesos que se desarrollan en la interacción requiere un acceso coordinado a los recursos que comparten. Los distintos procesos requerirán la consulta y actualización de la información almacenada en los diferentes modelos de conocimiento involucrados en la interacción. Tal es el caso del conocimiento sobre el estado de las metas compartidas, los usuarios, la sesión, la situación, las emociones, el conocimiento propio del sistema y la ontología.

Gestión de la Toma de Turno: Una toma de turno más compleja que el ciclo de la interacción requiere un conocimiento exhaustivo del estado en el que se encuentran los turnos que desarrollan los participantes de la interacción; conjeturar sobre qué participante recae la posesión de la palabra en cada momento; y estimar cuáles de ellos se perfilan como candidatos a tomarla. Toda esta información determina cuándo el sistema (en caso de requerirlo) podría tomar la palabra, cuándo la palabra recae sobre él y debe tomarla (aunque sólo sea para rechazarla), o cuándo se están produciendo silencios incómodos que conviene rellenar (de

acuerdo al conjunto de acuerdos tácitos que las personas aplican a sus propias interacciones [149]).

Gestión de Metas: Para soportar una toma de turno avanzada, el sistema deberá desempeñar un conjunto de nuevas funciones relacionadas con el tratamiento de las metas en la interacción. Entre ellas se encuentra recibir las metas discursivas que los diferentes modelos del sistema requieren desarrollar en la interacción; llevar a cabo su cancelación cuando el componente que las insertó decide retirarlas; y gestionar la forma en la que, tanto éstas como las comprometidas por todos los participantes, serán cursadas a través del diálogo.

Aunque existen algunas propuestas que abordan algunos de los aspectos de la *toma de turno avanzada* [Apartado 3.5], en ningún caso lo hacen de forma completa. Del mismo modo, todas ellas consideran al sistema un agente pasivo en lo referente a la toma de turno. Por tanto, se requieren soluciones que aborden, en conjunto, todos los problemas que comprende la toma de turno avanzada. Soluciones en las que el sistema (al igual que el resto de participantes) tome parte activa en el reparto de turnos de la interacción y en las que se considere la toma de turno, como a cualquiera de las actividades desarrolladas en la interacción [Apartado 2.2], una *actividad combinada*.

El presente trabajo pretende ofrecer una solución completa y de acción combinada a la toma de turno de la interacción natural. Para ello parte de la arquitectura de Sistema de Interacción Natural descrita en el Apartado 4.1 y la redefine y dota de componentes adicionales que permiten hacer frente a las nuevas habilidades interactivas descritas. Aunque el alcance de los cambios también afecta a los Componentes de Interfaz [Apartado 4.1.1] (para los que ahora se requiere que cumplan un conjunto de condiciones relacionadas con la granularidad con la que son capaces de adquirir y sintetizar las expresiones a través de las distintas modalidades), el núcleo de la propuesta son los componentes Gestor de Presentación y Gestor de Diálogo [Apartados 4.1.3 y 4.2.1]. Dichos componentes se ven afectados por la propuesta tal y como se describe [Figura 12]:

- *Gestor de Presentación*: Incorporará un nuevo componente, el Gestor de Turnos, que se articulará en los subcomponentes Gestor de Continuidad, Coordinador de Procesos y Gestor de Toma de Turno para soportar las funciones *gestión de continuidad*, *coordinación de procesos* y *gestión de la toma de turno*, respectivamente.
- *Gestor de Diálogo*: Incluirá el nuevo componente Gestor de Metas responsable de detectar cuándo las metas requieren ser desarrolladas por el sistema (tanto las

suyas propias como las combinadas) y de evaluar qué metas están en disposición de ser desarrolladas cuando se espere del sistema que intervenga en la interacción. Además su modelado deberá estar basado en las *teorías de la acción combinada* [Apartado 2.2] y soportar la distinción entre *metas colaterales y primarias* [Apartado 2.2.3.2].

A continuación se describen en mayor profundidad cada uno de estos componentes, así como la forma en que se desarrollan los procesos de la interacción natural según esta propuesta. También cómo hacen posible la gestión de una toma de turno avanzada.

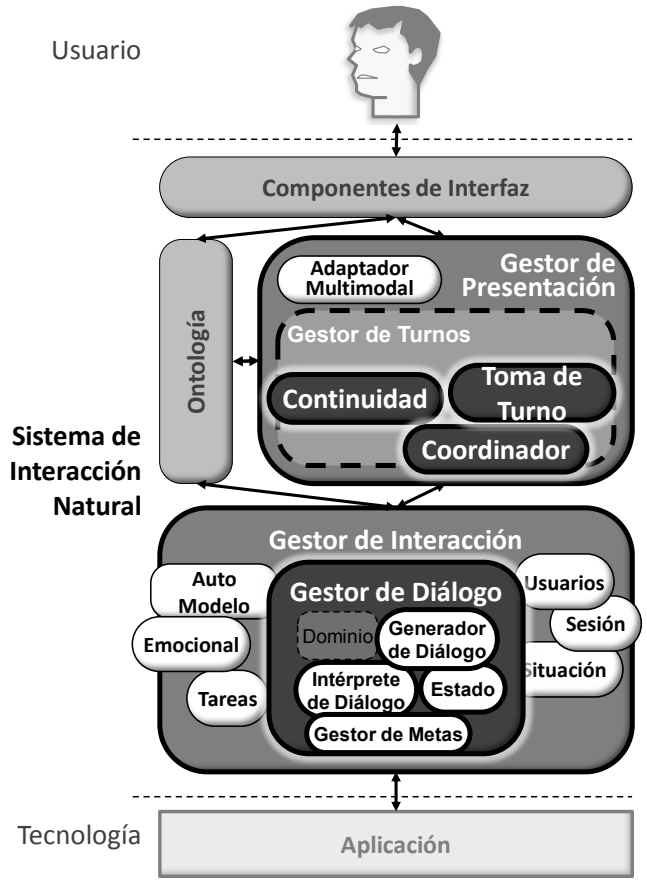


Figura 12: Arquitectura de Sistema de Interacción Natural para una *toma de turno avanzada*. Se resaltan los componentes añadidos o rediseñados por propuesta

5.1 *GESTOR DE TURNOS*

Ante un sistema de toma de turnos libre, cualquier contribución es construida conjuntamente por todos los participantes, tanto por el hablante como por sus oyentes, quienes a

través de sus contribuciones de realimentación ofrecen al hablante las evidencias de cierre a sus acciones que le permiten dirigir el curso de su intervención [Apartado 2.2]. Del mismo modo, los oyentes pueden aspirar por iniciativa propia a arrebatar al hablante la palabra, alterando o interrumpiendo su contribución, o simplemente indicándole que desean tomarla, quedando a su elección finalizar o no su contribución de forma prematura [Apartado 2.4]. De esta forma, ni la duración de las contribuciones, ni su orden o contenido están predefinidos de antemano. Surgen dinámicamente a medida que cada uno de los participantes evalúa su posición de cara a tomar o no la palabra y su necesidad, oportunidad y obligación de contribuir en la interacción (o de continuar haciéndolo). El resultado es una toma de turno no marcada, en la que tienen cabida incluso las situaciones de solapamiento o interrupción. Bajo este planteamiento de desarrollo temporal, las contribuciones no pueden ser tratadas como unidades atómicas que son producidas en un instante puntual en el tiempo en el todo que permanece inmutable. Como consecuencia, debe ser considerado que las contribuciones tienen un desarrollo temporal, y que los cambios que se dan en las circunstancias sociolingüísticas durante su propio desarrollo influyen en la formalización final, pudiendo alterar lo que estaba previsto expresar, o incluso hacer necesaria la rectificación de lo ya dicho. También se hace preciso coordinar la forma en la que los distintos procesos que se desarrollan en la interacción de forma simultánea interactúan entre ellos y utilizan sus recursos compartidos. Y, del mismo modo, surge la necesidad de modelar el estado en el que se encuentran los turnos de los distintos participantes; sobre quién recae la posesión de la palabra; e identificar qué participantes se perfilan como candidatos a contribuir, puesto que de toda esta información también depende la decisión de toma de turno.

En definitiva, surgen nuevas necesidades de presentación relacionadas con la sustentación de una organización temporal más potente que la toma de turno por *ciclo de interacción* [Apartado 2.1]. De esta forma, el Gestor de Presentación queda articulado en los subcomponentes Adaptador Multimodal [Apartado 4.1.3.1], para desempeñar las funciones de fusión y fisión multimodal, y Gestor de Turnos, sobre quien recaerán las nuevas habilidades de *gestión de continuidad*, *coordinación de procesos* y *gestión de la toma de turno* que posibilitan la ruptura del *ciclo de interacción*. Estos nuevos componentes relacionados con la gestión de turnos serán descritos a continuación [Apartados 5.1.1, 5.1.2 y 5.1.3, respectivamente].

5.1.1 Gestor de Continuidad

El Gestor de Continuidad, en colaboración con los componentes Coordinador de Procesos [Aparatado 5.1.2], Gestor de Toma de Turno [Apartado 5.1.3] y Gestor de Diálogo [Apartado 4.2.1], aborda el desarrollo incremental de los procesos de interpretación y generación de contribuciones en *la interacción natural*.

Desde un enfoque incremental, la interpretación y generación son procesos con un desarrollo continuo en el tiempo y que requieren ser ejecutados por los participantes en tiempo real. Sin embargo, estos procesos (que transcurren en paralelo en el tiempo) utilizan unos recursos compartidos (el conocimiento relativo al estado de interacción, sesión, situación, usuarios, auto modelo y emociones) y, en consecuencia, se requiere un control de acceso a dichos recursos. Para cumplir la restricción de tiempo real, este control de acceso no puede simplificarse al bloqueo de un proceso mientras se desarrolla el otro (cuya duración viene dada por la duración del turno que procesa), y se supondrá que es posible discretizar dichos procesos en unidades menores.

Será posible descomponer los procesos continuos de la interacción natural en subprocesos discretos si es posible identificar el valor máximo de la granularidad temporal soportada por los participantes sin que pierdan la percepción de que la interacción se desarrolla en tiempo real. Para cada uno de los niveles a los que se realiza la interpretación y la generación, es posible identificar una granularidad de procesamiento [Figura 13]. La unidad de procesamiento en el Estado de Interacción es el mensaje; en lo que respecta a la gestión del diálogo el acto comunicativo; para el procesamiento de lenguaje natural (nivel lógico de los Componentes de Interfaz) el n-grama; y para la adquisición y síntesis (nivel físico de los Componentes de Interfaz) la granularidad vendrá dada por el máximo retardo de respuesta que los interlocutores toleran sin perder la percepción de que la interacción se desarrolla en tiempo real (parámetro que está directamente relacionado con la tolerancia de los usuarios a los retardos en la recepción de las evidencias de cierre a sus acciones).

De todos estos niveles de granularidad, la única que impone restricciones temporales es la que concierne al nivel físico de los Componentes de Interfaz, por lo que éste será el valor máximo de la granularidad temporal soportada para una interacción natural. Los estudios realizados sobre el canal telefónico para determinar la tolerancia de los usuarios a los retardos de transmisión fijan este tiempo entre 150 y 400 ms. [91].

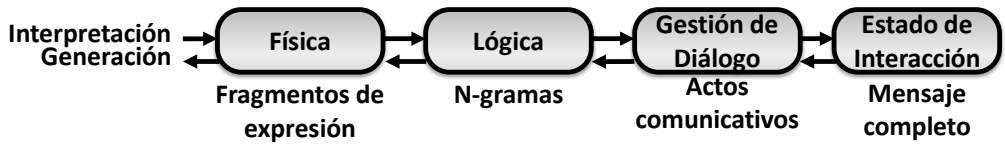


Figura 13: Granularidad de los procesos de interpretación y generación a los distintos niveles de la *interacción natural*

En definitiva, es posible descomponer los procesos de interpretación y generación de contribuciones, de naturaleza continua, en subprocesos discretos, siempre que estos no sean

ejecutados a intervalos mayores que la granularidad temporal máxima soportada. Se supondrá también que el tiempo de ejecución de los subprocesos puede ser considerado despreciable con respecto a la granularidad temporal definida (dado que los tiempos de cómputo de los ordenadores actuales son mucho menores que los tiempos de producción del lenguaje natural) y que, por tanto, no existirán retardos entre la solicitud y ejecución de un subproceso. De esta forma, el problema de la gestión de continuidad se restringe a la definición de dichos subprocesos y al establecimiento de su orden de ejecución en caso de colisión. Con ello quedan sentadas las bases para la definición una gestión incremental de los procesos de interpretación y la generación.

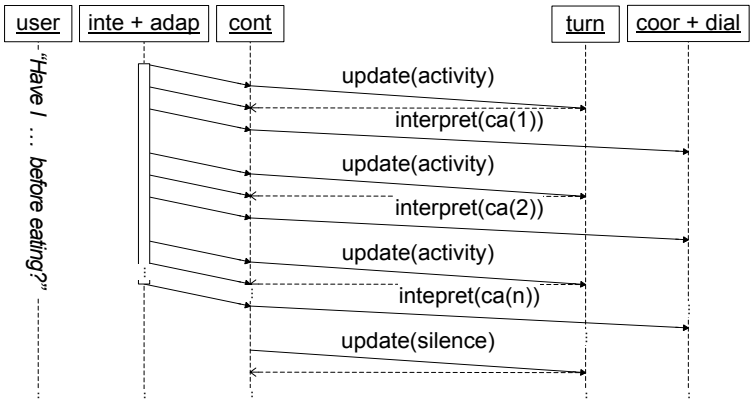
5.1.1.1 Interpretación Incremental

Las contribuciones de usuario, en lugar de ser interpretadas en un único proceso tras su completa adquisición, son interpretadas sobre la marcha a partir de una serie de subprocesos parciales. Las expresiones que realiza el usuario son adquiridas poco a poco a través de la interfaz y, a medida que el Gestor de Continuidad recibe los fragmentos de contribución del Adaptador Multimodal [Apartado 4.1.3.1], los combina con los anteriormente recibidos en unidades mayores para detectar porciones de contribución que, por si solas, representen acciones comunicativas completas. Cuando esto sucede, la porción de contribución recibida será considerada provisionalmente una contribución completa de su interlocutor y se solicitará al Gestor de Diálogo [Apartado 4.2.1] (a través del Coordinador de Procesos [Apartado 5.1.2]) la interpretación de los actos comunicativos que incluye, realizándose con ello los progresos oportunos en el estado de interacción y en el conocimiento sociolingüístico asociado. El Ejemplo 1 muestra cómo se desarrollaría bajo este planteamiento la interpretación de la contribución de usuario "*¿Tengo algo pendiente antes de comer?*".

Un enfoque incremental de la interpretación permite refinar iterativamente los actos comunicativos obtenidos a medida que se reciben nuevos fragmentos de contribución. En algunos casos, el resultado serán nuevos actos comunicativos con los que actualizar el estado de interacción y el resto de conocimiento sociolingüístico. En otros, denominados *reinterpretaciones*, supondrá incluso la rectificación de algunos de los actos comunicativos previamente interpretados por otros más precisos. En estos casos de reinterpretación, la interpretación de los nuevos actos comunicativos deberá estar precedida de una restauración previa del conocimiento de la interacción al estado anterior a la interpretación de los actos comunicativos obsoletos. Esta restauración podría incluso deshacer los cambios realizados por otros procesos de interpretación y generación, algunos de los cuales podrían haber desencadenado el desarrollo de metas que ahora quedarán canceladas. En estos casos será

posible desarrollar aclaraciones, disculpas o rectificaciones a través de la inserción de nuevas metas discursivas en el Gestor de Diálogo.

Del mismo modo, al ser recibidos nuevos fragmentos de contribución, el Gestor de Continuidad notifica la continuación de la actividad al Gestor de Toma de Turno [Apartado 5.1.3], o su cese (cuando se detectan silencios). También corresponde a este componente el reconocimiento de los marcadores relacionados con la gestión de la toma de turno que se expresan como alteraciones en la continuidad temporal de la contribución. Tal es el caso del *retardo momentáneo y reinicio del turno* como estrategia para tomar la palabra [71], o la *parada y continuación* (a través de carraspeos, repetición de las últimas palabras o cualquier otro tipo de silencios oralizados) para evitar perderla [92]. Todos estos marcadores serán remitidos, al igual que los actos comunicativos, al Gestor de Diálogo para su interpretación.



Ejemplo 9: Interpretación incremental de la contribución *"¿Tengo algo pendiente antes de comer?"*

Finalmente, la actualización del estado de interacción y resto de conocimiento sociolingüístico (tras la detección de nuevos actos comunicativos), y la actualización del estado de los turnos, la palabra y los candidatos (tras la detección del cese o continuación de la actividad y de los marcadores de toma de turno) son producidos durante el desarrollo de la contribución (y no sólo tras su completa adquisición), pudiendo afectar de forma simultánea al curso de la contribución actual del sistema, cancelarla o desencadenar la generación de una nueva.

5.1.1.2 Generación Incremental

Por efecto de los cambios que se producen en las circunstancias sociolingüísticas, de las contribuciones que producen el resto de participantes simultáneamente, o por la detección de errores en la interacción, la contribución que el sistema desarrolla en su turno actual puede

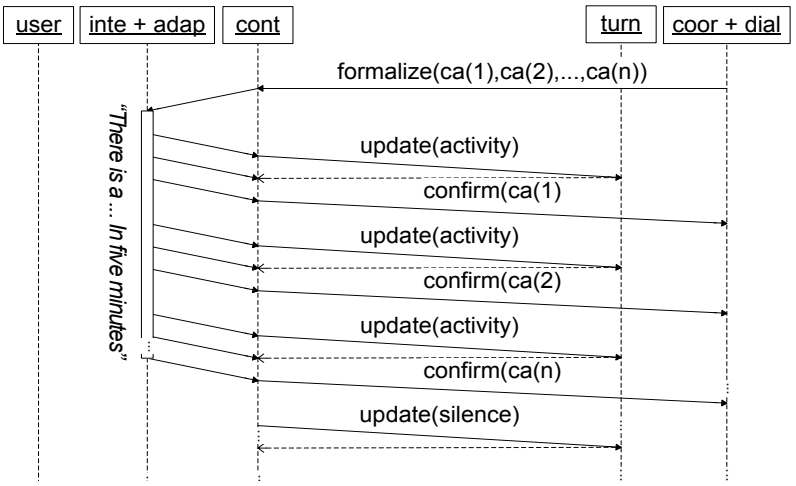
quedar obsoleta, haciendo necesaria su rectificación, reformulación o interrupción. Aunque la reformulación de una contribución es desarrollada por el Gestor de Diálogo [Apartado 4.2.1] y la decisión de la toma de turno proviene del Gestor de Toma de Turno [Apartado 5.1.3], la tarea de coordinar el proceso de generación incremental, adaptando los nuevos fragmentos de contribución del sistema a los ya expresados (incluso ante situaciones de rectificación o auto interrupción) es desarrollada por el Gestor de Continuidad.

Para gestionar la continuidad de la reformulación con la mayor fluidez posible, se requiere descomponer el proceso de generación de las contribuciones del sistema en dos subprocesos diferentes: su formalización y la confirmación de síntesis de las unidades que la componen. La formalización es la primera fase de la generación y se produce en el Gestor de Diálogo (bajo el control del Coordinador de Procesos [Apartado 5.1.2]). Es en este momento cuando se construye la contribución que el Gestor de Continuidad incorporará al flujo de expresión del sistema. Por su parte, la confirmación de síntesis consiste en notificar al Gestor de Diálogo cuándo han sido expresados fragmentos de contribución con sentido interactivo completo para actualizar los progresos que suponen en el estado de interacción y en el del resto de modelos de conocimiento (Modelo de Sesión, Modelo Emocional, etc.).

Esta propuesta considera que, en lo que respecta a la gestión del dialogo, las unidades atómicas en que el sistema estructura la formalización de su contribución son los actos comunicativos. Por ello, considera que la expresión completa de cada uno de estos actos comunicativos será el desencadenante una nueva confirmación en el Gestor de Diálogo. A través de este mecanismo, el Gestor de Continuidad podrá conocer qué porción de la contribución ha expresado el sistema hasta el momento en su turno y qué parte, en su caso, queda por expresar. Identificar la porción de contribución que ha sido expresada hasta el momento es fundamental para hacer posible una posterior gestión de la reformulación. El Ejemplo 10 muestra como se desarrollaría la generación incremental de la contribución del sistema "*Tiene una reunión en cinco minutos*".

La reformulación se produce cuando, teniendo el sistema una contribución en curso, se produce una nueva decisión de toma de turno [Apartado 5.1.3.4] que resulta en una nueva formalización. En estos casos, se considerará que existe un *punto de transición* entre las formulaciones si la nueva comienza por un *acto comunicativo compatible* con alguno de los actos comunicativos de la formalización previa. En función de que exista o no un punto de transición entre las contribuciones (marcado como "|") y de dónde se encuentre podrán distinguirse los casos de *reformulación suave*, *rectificación* y *auto interrupción*. La reformulación suave se produce cuando ambas formalizaciones, obsoleta y nueva, encajan y el

punto de transición es posterior a la porción expresada hasta el momento ("*Tiene una reunión en|... cuatro minutos"*). Por su parte, la rectificación ocurre cuando ambas formalizaciones encajan, pero el punto de transición ya había sido expresado ("*Tiene una reunión en cinco min|... en cuatro minutos*"). El resultado será una *auto interrupción* cuando las formalizaciones no encajen de ningún modo ("*Tiene una reunión en cinco min|... disculpe, tiene una llamada*", Ejemplo 11).



Ejemplo 10: Generación incremental de la contribución "*Tiene una reunión en cinco minutos*"

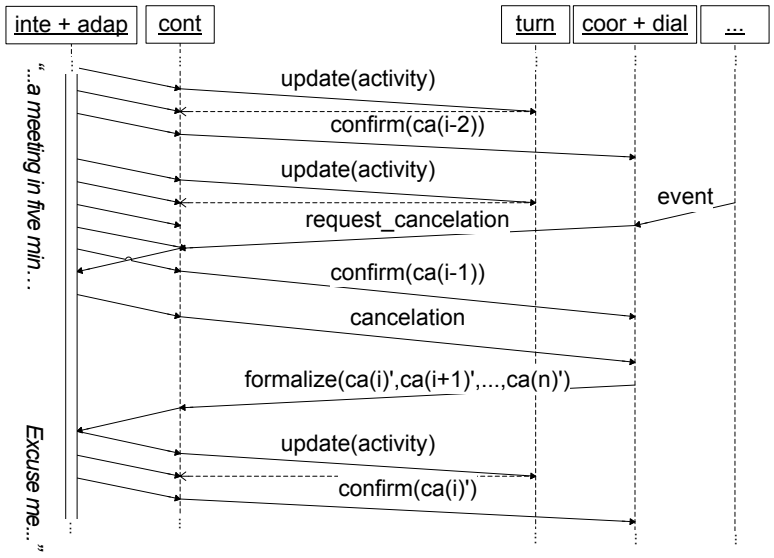
Finalmente, entre las funciones del Gestor de Continuidad también se encuentra la de notificar al Gestor de Toma de Turno cuándo hay actividad en el turno del sistema o cuando cesa (desarrolladas durante los procesos de confirmación). También la de sintetizar, en forma de alternaciones en la continuidad temporal del turno del sistema, algunos marcadores de gestión de turno (*retardo momentáneo del turno*, *parada y continuación*, etc. [32, pp. 266-274]) cuando así esté definido en la formalización recibida del Gestor de Diálogo.

Con todo ello, el proceso de generación no queda restringido a la expresión de lo formalizado inicialmente. El sistema puede actualizar su contribución sobre la marcha para ajustarla a los cambios que se producen simultáneamente en el estado de interacción; en el estado de los turnos; en la posesión de la palabra; en el estado de los candidatos a tomarla; o en las circunstancias sociolingüísticas.

5.1.2 Coordinador de Procesos

Durante el desarrollo de la *interacción natural*, existen varios procesos que se desarrollan simultáneamente en el sistema. Por un lado, la interpretación de las contribuciones de los interlocutores, que va siendo realizada de forma continúa a lo largo del tiempo. Por otro,

la generación de las propias contribuciones que el sistema puede desarrollar en paralelo a estos procesos de interpretación. Finalmente, las circunstancias que rodean a la interacción van cambiando dinámicamente y se requiere monitorizar los cambios que se producen en ellas para detectar cuándo es preciso desencadenar nuevos procesos de generación del sistema.

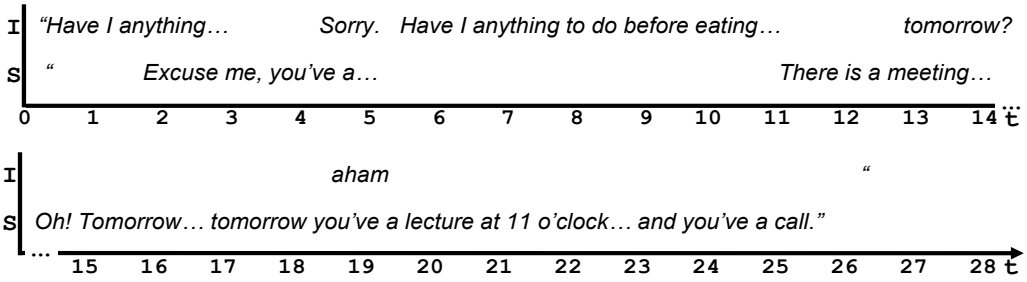


Ejemplo 11: Reformulación de la contribución "*Tiene una reunión en cinco minutos*"

Algunas situaciones en las que concurren varios procesos son los representados en el Ejemplo 12. Se dan procesos simultáneos cuando, por ejemplo, alguno de los participantes ofrece realimentación simultánea al hablante (*t=19*) o cuando se dan solicitudes de posesión de la palabra (*t=2*), etc. También cuando, bien por error en la estimación del estado de los turnos, la posesión de la palabra o los candidatos (*t=11*), o por competir por la posesión de la palabra (*t=4*), los participantes intervienen de forma solapada, pudiendo incluso interrumpirse unos a otros (*t=5*). Además, diversos eventos (con origen en cambios en las circunstancias sociolingüísticas, el estado de la interacción o en la propia pro actividad del sistema) podrían requerir la revisión del turno del sistema (toma o cancelación de turno o la reformulación de la contribución en curso). En el Ejemplo 13 se observa la pérdida de eficiencia y naturalidad que conlleva limitar la capacidad del sistema para desarrollar en paralelo dichos procesos, por lo que se hace necesario desarrollar estrategias de coordinación que permitan atenderlos de forma concurrente.

Con el tratamiento incremental de la interpretación y generación de contribuciones, estos procesos (de naturaleza continua) quedan descompuestos en subprocesos discretos de tiempo de ejecución despreciable frente a los tiempos de producción del lenguaje natural. Estos

procesos son: *interpretación de actos comunicativos*, *formalización de contribuciones* y *confirmación de expresión de actos comunicativos*.

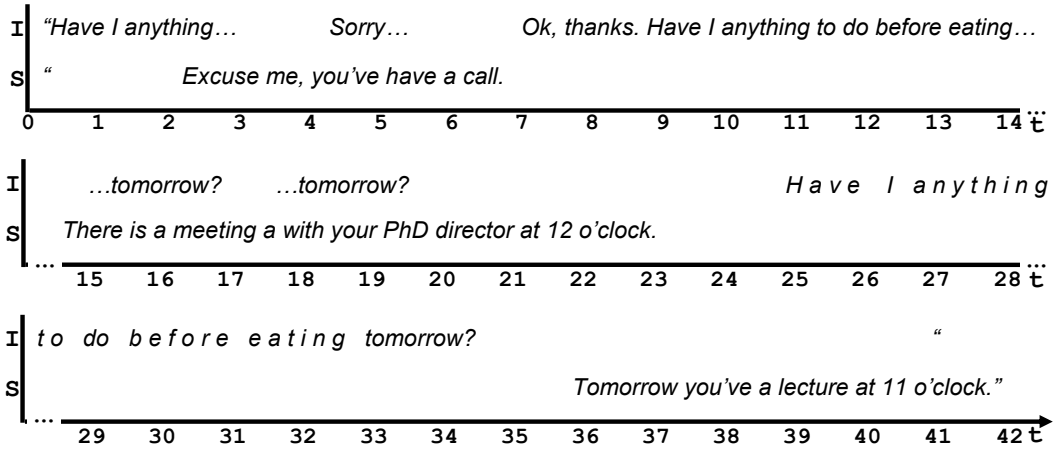


Ejemplo 12: Diálogo con desarrollo simultáneo de procesos

La interpretación de actos comunicativos es desencadenada por el Gestor de Continuidad [Apartado 5.1.1] cuando son detectados en la contribución en curso de algún interlocutor nuevos fragmentos con sentido interactivo completo (actos comunicativos). Por su parte, la formalización de contribuciones es desencadenada cuando el Gestor de Toma de Turno detecta cambios en el estado de los turnos; en la posesión de la palabra; o en los candidatos a tomarla, y estos pudieran suponer una toma o cesión de turno del sistema (o la reformulación de su contribución en curso). También cuando el Gestor de Metas [Apartado 5.2.2] lo hace con respecto a la criticidad de las metas discursivas propias del sistema o el compromiso de las metas combinadas. Las confirmaciones de expresión de actos comunicativos son solicitadas cuando, desde los Componentes de Interfaz de Salida [Apartado 4.1.1.2], se notifica que han sido sintetizados en el canal todos los fragmentos de contribución que comprenden la expresión de cada uno de los actos comunicativos de la componen. Dado que cualquiera de estos subprocesos requiere el acceso a una serie de recursos compartidos (el conocimiento sobre el estado de interacción, la sesión, usuario, emociones, situación y auto modelo) y que el orden en el que se ejecutan es determinante en el estado final alcanzado, la gestión de la continuidad debe ser complementada con una estrategia de coordinación de procesos.

Los progresos que produce la confirmación de cada acto comunicativo sobre el Estado de Interacción y el conocimiento sociolingüístico asociado [apartados 5.2.1 y 4.2.3, respectivamente] sólo tienen sentido con respecto al estado alcanzado por la confirmación de los actos comunicativos previos. En el caso del primer acto comunicativo de una contribución, con respecto al estado en el que fue formalizada la contribución. Por ello, las confirmaciones de expresión de los actos comunicativos sólo tienen sentido si se producen a continuación de la formalización de la contribución a la que pertenecen (sin interponerse ningún otro proceso que pudiera alterar el conocimiento compartido) y en el mismo orden en el que figuran en la contribución. Debido a la concurrencia de procesos que tiene lugar en la interacción natural, en

ocasiones pueden interponerse, entre la formalización de una contribución y la confirmación de la expresión de sus actos comunicativos, otras solicitudes de servicios (interpretaciones de actos comunicativos de usuario o formalizaciones de nuevas contribuciones de sistema). De ser atendidas estas solicitudes previamente a la confirmación de los actos comunicativos que ya han sido expresados, la confirmación se produciría sobre un estado mutado en el que la expresión de dicho acto comunicativo podría haber perdido su sentido. Sin embargo, que el acto comunicativo se expresó es un hecho conocido por todos los participantes (puesto que su confirmación de síntesis fue recibida), por lo que su confirmación deberá preceder a la ejecución de cualquier otro tipo de solicitud. De esta forma, la confirmación sobre el estado de interacción de los actos comunicativos de los que se tiene constancia que fueron expresados constituye, por tanto, el tipo de solicitud de mayor prioridad en lo que a la coordinación de procesos se refiere.



Ejemplo 13: Diálogo con desarrollo secuencial de procesos

Dado que la atención de estas nuevas solicitudes (interpretación de nuevos actos comunicativos de los interlocutores y formalización de nuevas contribuciones del sistema) supondrá cambios en el estado de conocimiento que podrían causar la pérdida de vigencia de la contribución en curso, será preciso gestionar la reformulación de la contribución en curso cuando éstas se produzcan. Para ello, en primer lugar, será cancelada la expresión de la porción pendiente de contribución (la correspondiente a los actos comunicativos cuya expresión había sido notificada hasta el momento). Tras ello será ejecutada la nueva solicitud. Si consiste en una solicitud de formalización, el resultado será la reformulación de la porción de contribución que fue cancelada. Se distinguirá entre una reformulación suave, rectificación o auto interrupción de la contribución en curso en función del alcance de los cambios que desencadenaron dicha solicitud. Si se trata de la solicitud de interpretación de nuevos actos comunicativos de alguno de los interlocutores del sistema y, tras su ejecución, el estado de las metas y el de la toma de

turno siguen justificando el desarrollo de la porción de contribución que fue cancelada (o parte de ella), el Gestor de Metas o el Gestor de Toma de Turno desencadenará una nueva solicitud de formalización. En este caso, la porción pendiente de contribución será reformulada sobre los cambios ocurridos en el estado de conocimiento como consecuencia del proceso de interpretación que se interpuso.

En cuanto a la relación de prioridad entre los procesos de interpretación de actos comunicativos y los de formalización de contribuciones, debe considerarse que la interpretación se refiere al progreso que producen en el estado de conocimiento acciones que ya han sido realizadas por los interlocutores (puesto que han sido recibidas), por lo que su prioridad es mayor a la formalización de una nueva contribución.

De esta forma, el orden de prioridad en la tramitación de las distintas solicitudes será (de mayor a menor):

1. Confirmación de la expresión de actos comunicativos
2. Interpretación de actos comunicativos
3. Formalización de nuevas contribuciones

Aunque este reparto de prioridades permite garantizar la consistencia en el estado de información alcanzado por el sistema, en ocasiones resulta imposible garantizar una sintonía entre su representación de la interacción y la que mantienen otros participantes. Esta situación podrá darse ante la concurrencia de solicitudes de interpretación de actos comunicativos de distintos interlocutores o ante la concurrencia de una solicitud de interpretación de actos de un interlocutor y la de confirmación de expresión de actos comunicativos del sistema (aunque sólo cuando las solicitudes que concurren afecten a un conocimiento común).

Bajo estos supuestos y, a pesar que una confirmación debe ser procesada previamente a una interpretación para garantizar la consistencia en el estado de interacción alcanzado por el sistema, podría darse que ambos actos comunicativos partiesen del mismo estado inicial para alcanzar estados incompatibles. Para detectar estas situaciones, el sistema debe analizar el estado de conocimiento que se alcanzaría de invertir el orden de los procesos. Cuando existe una divergencia entre el estado que alcanza y el que alcanzaría siguiendo el otro orden, existe un riesgo de pérdida de la sintonía entre las representaciones de la interacción que mantienen los distintos participantes (puesto que no puede saberse con seguridad en qué orden los ejecutaron otros participantes). En estos casos, tal y como sucede en la interacción desarrollada entre personas, los participantes penalizarán el compromiso de los hilos desarrollados en función de

la divergencia encontrada. Si la caída del compromiso es suficientemente grave, es de esperar que esto desencadene reparaciones en la interacción, como reformulaciones de las contribuciones en curso o interrupciones.

5.1.3 Gestor de Toma de Turno

La interacción basada en el *ciclo de interacción* establece unas reglas rígidas de alternancia de turno bajo las cuales sólo el hablante puede producir turno. Además, la posesión de la palabra es pasada de un participante a otro a lo largo del desarrollo de toda la interacción y es el hablante quien estructura unilateralmente su turno, en duración y contenido. En realidad, tal y como reflejan los estudios sobre el *uso del lenguaje* [Apartado 2.4], tomar turno en la interacción natural conlleva una mayor dificultad que todo eso.

En la gestión de la toma de turno entran en juego el estado de la interacción [Apartado, 4.2.1.1] y las circunstancias sociolingüísticas que la rodean (información ya representada por otros modelos de conocimiento [Apartado 4.2.3]). Además de todo este conocimiento, también es preciso considerar el estado en el que se encuentran los turnos que están desarrollando los participantes, quién está en posesión de la palabra y qué participantes se perfilan como candidatos a tomarla. Representar esta información será una de las funciones del Gestor de Toma de Turno [Apartados 5.1.3.1, 5.1.3.2 y 5.1.3.3].

Junto con esta función, el Gestor de Toma de Turno será también responsable de elaborar las decisiones de toma de turno en base a las cuales el sistema determina, desde un planteamiento de actividad combinada, los momentos en los que tiene cabida que el sistema genere contribuciones para hacer progresar algunas de las metas [Apartado 5.1.3.4]. La decisión de la toma de turno se desencadenará ante los cambios ocurridos en el estado de conocimiento, en cualquiera de los aspectos anteriormente mencionados, y conlleva la revisión de la urgencia de cada una de las metas interactivas. En función de las circunstancias en las que esto se produzca, podrá conllevar la formalización de una nueva contribución del sistema, la reformulación de su contribución en curso, o su auto-interrupción.

5.1.3.1 Estado de Turnos

A lo largo de la propuesta se ha aplicado el término *contribución* en referencia a las aportaciones que hacen los participantes en la interacción. Esto es, a la sucesión de expresiones producidas a través de distintas modalidades que representan las acciones comunicativas con las que los participantes pretenden hacer progresar la interacción a nivel local y global. Por su parte, el término *turno* hace referencia al hecho de ejecutar temporalmente una contribución,

más allá de su propio contenido. Turno, como tal, no ofrece información de las metas que se pretende desarrollar, ni de los progresos que se producen en ellas, ni de cómo se altera el conocimiento sociolingüístico involucrado (cuestiones que serán tratadas durante la interpretación y generación de diálogo [apartados 5.2.3 y 5.2.4, respectivamente]).

El sistema modelará el estado en el que se encuentra el turno de cada uno de los participantes de la interacción (incluyendo el suyo propio). El estado del turno de un participante refleja si está desarrollando actividad en el canal, si se ha alcanzado un posible fin de la contribución (*transition relevant place* o *TRP* [154, pp.3-12]), y si las metas desarrolladas por el turno son colaterales o primarias a los denominados *asuntos oficiales* de la interacción. Dado que esta gestión es una actividad combinada, se hace desde el punto de vista de la conjetura que podrían alcanzar los interlocutores sobre este estado del turno, y no como representación de la actividad individual realizada por un participante. Estas conjeturas, como cualquier otra referente al conocimiento mutuo, podrán debilitarse y requerir los pertinentes refuerzos y reparaciones.

El estado de actividad de los turnos y la estimación sobre el fin de su contribución es actualizado por el Gestor de Continuidad y el Gestor de Diálogo [apartados 5.1.1 y 4.2.1] durante los procesos de interpretación y generación (conjetura el estado si el turno pertenece a un interlocutor y lo actualiza cuando es del sistema). El Gestor de Continuidad actualiza el estado del turno cuando se detectan nuevos fragmentos de contribución (*actividad*), cuando se producen *silencios* y cuando se alcanzan porciones de contribución con sentido interactivo completo (*actos comunicativos*). Del mismo modo, la gestión de continuidad también permite deducir situaciones de actividad que no hacen progresar la interacción, sino que constituyen recursos, como los titubeos o la repetición de palabras, que son producidos con el objetivo de rellenar silencios (*silencios oralizados*). Por su parte, el Gestor de Diálogo actualiza el estado de actividad de los turnos y la estimación sobre el posible fin de la contribución cuando se alcanzan explícita o implícitamente *TRPs*. Por ejemplo, una ocurrencia explícita puede expresarse verbalmente ("*¿Tú qué opinas?*") o mediante gestos específicos para indicar que se ha terminado. Por otro lado, las ocurrencias implícitas se deducirán del estado de interacción alcanzado (por ejemplo, cuando el interlocutor ha finalizado una transacción, sus metas pendientes, o ha llegado a un punto crítico de la interacción).

El carácter primario o secundario del turno que desarrolla un participante será parte de la información recogida en el estado de su turno. A esto se le conocerá con el nombre de pista de acción del turno de un participante. La conjetura acerca de la pista de acción de un turno define si éste versa sobre los asuntos oficiales de la interacción (turno primario) o sobre

cuestiones colaterales a ellos (turno secundario). La conjetura de la pista se realiza en el Gestor de Diálogo durante el proceso de generación o interpretación (según sea turno del sistema o de otro participante) y representa la relación primaria o secundaria que existe entre el hilo que actualmente desarrolla el turno y el que se considera foco actual de la interacción. La pista de un turno puede cambiar a medida que se suceden los distintos discursos de la contribución o cuando se actualiza en la *estructura focal* [Apartado 4.2.1.1] el *hilo enfocado*.

5.1.3.2 Estado de Posesión de la Palabra

El Gestor de Toma de Turno también representa sobre qué interlocutor recae la palabra o, dicho de otro modo, quién es el hablante actual. Esta estimación es realizada a partir de la información relativa al estado de los turnos.

Partiendo de los supuestos más comúnmente aceptados sobre las reglas que rigen la toma de turno humana [Apartado 2.4] se considera que, cuando la palabra está vacante (no hay ningún hablante en curso), el primer participante en tomar el turno será considerado hablante. Del mismo modo, en cada instante sólo podrá haber un hablante, aun cuando varios participantes estuvieran interviniendo simultáneamente (incluso con carácter primario). El hablante será el primero que comenzó a intervenir, y sólo tras haber desistido (cuando su turno queda inactivo) o cuando ha expresado una TRP o pudiera interpretarse que la expresó, la palabra pasará a otro participante, siguiendo el orden de prioridad descrito en dichas reglas de toma de turno.

Cuando transcurre un tiempo demasiado largo sin actividad por parte del hablante (ni siquiera para rechazar su turno) y no existe ningún otro participante que haya aspirado a tomar la palabra, ésta quedará vacante (produciendo silencios incómodos que conviene rellenar). Ante tales situaciones el Gestor de Toma de Turno desencadenará nuevas solicitudes de formulación de contribución que depositará en el componente Coordinador de Procesos [Apartado 5.1.2]. El resultado podrá ser la formalización de turnos secundarios (como gestos) para retener la palabra cuando se supone que el sistema está en posesión de la palabra y está interesado en seguir estándolo (por ejemplo para continuar desarrollando una contribución cuya formalización requiere demasiado tiempo), o para tomar la palabra cuando está interesado en hacerlo y el hablante está prolongando injustificadamente su turno.

5.1.3.3 Participantes Candidatos a Tomar la Palabra

De cara a tomar el turno para desarrollar nuevas contribuciones, especialmente aquellas que ocupan la pista primaria, o para decidir si la palabra se puede retener, es fundamental

conocer la posición de los participantes frente a turnos futuros. Esto implica conocer si los participantes han sido apuntados como posibles hablantes siguientes (a través de gestos, vocativos, alusiones o, simplemente, porque dependa de ellos el desarrollo del foco) o si han solicitado la palabra (a través de gestos, estrategias de ganancia de la atención, etc.). Para estimar la posición de los participantes frente a turnos futuros entran en juego la organización local y global de la interacción, las técnicas aplicadas sobre la continuidad de la contribución y la evolución de las circunstancias sociolingüísticas en las que se desarrolla. Por ello, la posición de los participantes frente a turnos futuros será actualizada durante los procesos de generación e interpretación tanto por el Gestor de Continuidad como por el Gestor de Diálogo [apartados 5.1.1 y 4.2.1].

5.1.3.4 La Decisión de la Toma de Turno

Los cambios en el estado de la interacción y en el conocimiento sociolingüístico asociado (como consecuencia de las contribuciones que los participantes hacen, o por cambios en las circunstancias en las que se desarrolla la interacción), alterarán la criticidad de las metas discursivas propias del sistema, el compromiso de los hilos combinados y el estado de la toma de turno. Del mismo modo, cualquiera de los componentes de la arquitectura tiene capacidad para insertar nuevas metas en el estado de interacción (de forma autónoma y en cualquier momento) si así se requiere para el desarrollo de sus propias tareas. Esto ocurrirá, por ejemplo, cuando el Gestor de Toma de Turnos [Apartado 5.1.3] dude a la hora de conjeturar si el sistema debe continuar en posesión de la palabra ("*¿Sigo?*"), o cuando la Ontología [Apartado 4.1.2] encuentre problemas para identificar el concepto al que se refiere algún término que utilice su interlocutor ("*¿Qué quieres decir con…?*"). Por ello, ante cualquiera de estos cambios, el sistema deberá analizar si las circunstancias hacen posible que el sistema tome, retenga o libere el turno.

La decisión de toma de turno se desarrolla en el Gestor de Toma de Turno y consiste en la evaluación de la urgencia con la que el sistema debe desarrollar cada una de las metas de la interacción. La urgencia de cada meta es calculada a partir de su criticidad, del compromiso alcanzado en su hilo combinado y de la expectativa de beneficio con respecto al coste [102]. Estos valores de urgencia permitirán al componente Generador de Diálogo determinar la medida en que las metas de la interacción para las que el sistema tiene posibles progresos están en condiciones de ser desarrolladas (dado el estado actual de los turnos, la palabra y de quienes sean los candidatos a tomarla).

5.2 *GESTOR DE DIÁLOGO PARA UNA TOMA DE TURNO AVANZADA*

Aunque éste no es un componente con competencias directas sobre la organización temporal de la interacción, sus funciones afectan directamente a la toma de turno. De la organización local de cada una de las metas compartidas se deduce en qué ocasiones su desarrollo depende del sistema (posicionándole como candidato a tomar turno); de la caída de su compromiso posibles contribuciones de realimentación o incluso interrupciones encaminadas a mejorarlo; de las relaciones de dependencia entre unas metas y otras cuándo es buen momento para retomar o proponer nuevas "ideas" sobre las que el sistema podría tener un especial interés (pudiéndole hacer solicitar o arrebatar la palabra al hablante en curso); de lo crítico que resulte para el sistema desarrollar sus propias metas discursivas el que compita para hacerse con el poder sobre la palabra, etc.

Por todo ello, este componente es, junto con el Gestor de Presentación [Apartado 4.1.3], núcleo de la propuesta. En este apartado se revisará la forma en que deben ser rediseñados los componentes Estado de Interacción, Intérprete de Diálogo y Generador de Diálogo [apartados 5.2.1, 5.2.2 y 5.2.4, respectivamente] para adecuarlos a las necesidades de una toma de turno avanzada. Del mismo modo, será descrito el nuevo componente Gestor de Metas [Apartado 5.2.2], responsable de la monitorización de la evolución del compromiso alcanzado sobre las metas combinadas y de la criticidad de las metas discursivas propias del sistema.

5.2.1 Estado de Interacción

El Estado de Interacción representa la interacción a los niveles local y global. De esta forma contiene la información relativa al estado en el que el sistema supone que se encuentra cada uno de los hilos comprometidos por los participantes, sus posibles desarrollos futuros, su "salud" (el valor del compromiso respecto a la atención, interés e información), su dependencia jerárquica de otros hilos (organización intencional) y el orden en que éstos se desarrollan (organización focal). Todos estos aspectos influyen en la toma de turnos, bien por caídas en el compromiso de alguno de los hilos desarrollados, o bien por la forma en la que el estado de interacción alcanzado puede apuntar a determinados participantes como candidatos a tomar turno en la interacción.

Por otro lado, la toma de turno avanzada requiere la capacidad de proyectar prematuramente los progresos que supondrá sobre el estado de interacción una contribución de sistema en el momento de su formalización (y que posteriormente irán siendo confirmados al notificarse la expresión de cada uno de los actos comunicativos que la componen) y la de

recuperar estados de interacción anteriores para desarrollar las reinterpretaciones (al progresar la contribución del interlocutor y ser posible obtener actos comunicativos más precisos o al ser detectados errores en la interpretación previa). De esta forma, se requiere un estado de interacción con control de versiones.

Finalmente, es también necesario mantener conjeturas sobre la *pista de acción* [Apartado 2.2.3.2] en la que se desarrolla el turno de cada participantes de la interacción, por su importancia en la decisión de la toma de turno. De esta forma, se precisa modelar la pista de acción que cada uno de los hilos abiertos de la interacción lleva asociada.

5.2.1.1 Gestión del Compromiso

La evolución del compromiso durante los proceso de interpretación y generación de diálogo influirán en la toma de turno desarrollada en la interacción. A medida que el compromiso alcanzado en alguno de los hilos se degrade, el sistema (al igual que hacen el resto de participantes) requerirá desarrollar estrategias de cara a restaurar la salud del hilo. Las estrategias de reparación del compromiso tratarán de ser aplicadas sin agredir al desarrollo de la intervención primaria en curso. Sin embargo cuando las caídas son pronunciadas, se requerirán estrategias más agresivas, que pueden suponer solapamientos (e incluso interrupciones).

Aunque es el componente Gestor de Metas [Apartado 5.2.2] quien se encarga de monitorizar el estado de los hilos combinados y, en consecuencia, será él quien solicitará la formalización de contribuciones al Generador de Diálogo [Apartado 5.2.4] a través del Coordinador de Procesos [Apartado 5.1.2] ante las caídas en sus compromisos, la representación de dicho parámetro recae sobre el Estado de Interacción.

5.2.1.2 Estimación de los Candidatos a Tomar la Palabra

Dejando aparte las designaciones de participantes y las solicitudes de palabra realizadas expresamente (a través del desarrollo de juegos de diálogo específicos), el estado de interacción permite estimar aquellas situaciones en las que, indirectamente, alguno de los participantes se presenta como un posible candidato de cara a la toma de la palabra.

Los hilos describen los posibles progresos futuros que cabe esperar del desarrollo de las metas que representan. Por ello, en cada uno de sus estados, es posible determinar de qué participantes depende su desarrollo y, en caso de tratarse del hilo enfocado, tales hablantes podrán considerarse participantes designados (indirectamente) para tomar la palabra. Cuando el desarrollo del foco requiere el desarrollo previo de otros hilos, aquellos participantes de los que

dependan sus progresos podrán ser también considerados participantes designados a tomar la palabra.

Por otro lado, el haber alcanzado estados que requieren contribuciones de otros participantes (bien en el foco o en los hilos de los que dependa su desarrollo) ayudan a estimar con mayor certeza aquellos *TRPs* que no son realizados por acciones específicas por parte del hablante, sino que deben interpretarse de los silencios posteriores a la terminación de acciones comunicativas completas.

5.2.1.3 Control de Versiones

Se considera que todo acto comunicativo desencadena una actualización en el estado de interacción y en el conocimiento sociolingüístico asociado. Cada vez que puede darse por expresado un acto comunicativo en el canal (sea de usuario o de sistema), los participantes actualizan su estado (desde el que estiman como actual hasta el más probable de todos aquellos que pueden ser alcanzados con dicho acto comunicativo). Cuando se trata de la interpretación de las intervenciones de los interlocutores, poco a poco se irán recibiendo nuevos actos comunicativos (o la reinterpretación de otros previos) y, con cada uno de ellos, se alcanzará un nuevo estado en el conocimiento compartido. Desde el punto de vista de la generación, cada vez que pueda darse por expresada la porción de contribución que se corresponde con cada acto comunicativo, podrán ser confirmados los progresos que se definieron durante la formalización para dicho acto comunicativo.

Sin embargo, el desarrollo de la *interacción natural*, especialmente cuando se produce bajo una toma de turnos no marcada, no permite garantizar que las conjeturas realizadas por los distintos participantes sobre el estado actual sean iguales. Por ello, en ocasiones, los participantes deben desarrollar estrategias encaminadas a hacer comparables las creencias que todos mantienen sobre la interacción. En algunos casos, la causa puede estar en las distintas posibles interpretaciones de un mismo acto comunicativo. En otros, errores cometidos en alguna de las fases de la generación, la interpretación o por problemas en la transmisión. Del mismo modo, una variación en el orden en que los distintos participantes desarrollan los subprocesos de que consta la interacción (interpretación de actos comunicativos en el diálogo, formalización de nuevas contribuciones de sistema y confirmación de la expresión de actos comunicativos) pueden llevar a estados diferentes. La resolución de tales problemas precisa de la capacidad de deshacer los progresos realizados de forma errónea sobre el estado de interacción y sobre su conocimiento sociolingüístico asociado. Con ello se hace posible recuperar estados anteriores

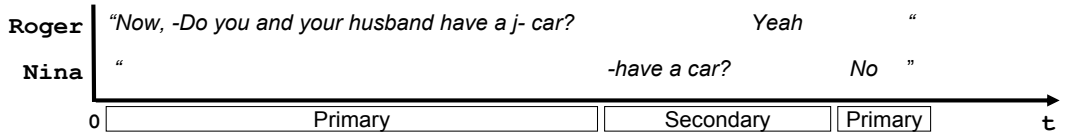
no erróneos, a partir de los cuales rehacer los progresos de la interacción alcanzando un mayor compromiso entre los participantes (por ejemplo en los casos de reinterpretación).

En consecuencia, el estado de interacción, al igual que su conocimiento asociado debe soportar el restablecimiento de versiones anteriores y la proyección de versiones futuras aún no confirmadas. La unidad de progreso en el estado de interacción es el acto comunicativo y con cada uno de ellos será registrada una nueva versión del estado de interacción.

5.2.1.4 Representación de Pistas

Tal y como se explicó anteriormente, la correcta representación del estado de los turnos requiere conocer qué pista desarrolla cada uno de ellos, por lo que es necesario modelar la estructura de *pistas de acción* [Apartado 2.2.3.2] de los hilos que se desarrollan la interacción. Esta función recae sobre el Estado de la Interacción.

Los estudios teóricos sobre la comunicación colateral en la *interacción humana* [Apartado 2.2.3.2] clasifican los asuntos que pueden ser desarrollados por los participantes como *primarios* o *secundarios* con respecto a los "temas oficiales" tratados en la interacción. Si desarrollan "ideas" que encajan dentro de aquellas que son consideradas como los asuntos que desarrollan "oficialmente" los participantes en un determinado momento ("*Now, -Do you and your husband have a j-car?*" [Ejemplo 14]), éstos se considerarán primarios. Sin embargo, si desarrollan ideas metacomunicativas con respecto a los temas oficiales ("*-have a car?*") serán considerados secundarios. Del mismo modo, todo asunto secundario puede ser, a la vez, primario con respecto a otros posibles asuntos que pudieran desarrollarse posteriormente en la interacción.



Ejemplo 14: Ejemplo de *pistas de acción* [Ref. Clark, 96]

Por tanto, el Estado de Interacción mantendrá una representación jerárquica de las pistas de acción desarrolladas en la interacción. Cada pista será secundaria para su pista padre y cualquier ancestro suyo y, a su vez, será primaria para sus hijos y todos sus descendientes. Cada uno de los hilos desarrollados en la interacción llevará asociada una pista, pudiendo esta ser la misma que la del hilo padre o una secundaria. Así, el hilo base desarrollará siempre la pista raíz de la interacción (puesto que representa a la propia conversación y, por tanto, asuntos oficiales) y el resto de hilos, en función de las condiciones de su apertura (descritas por el Dominio de

Interacción [Apartado 4.2.1.1]), podrán desarrollar la misma pista que su padre o una secundaria a ella.

En cualquier instante, un hilo será considerado primario (que desarrolla "asuntos oficiales") cuando tenga asociada la misma pista que el hilo actualmente enfocado, o una más primaria (alguna de sus ancestros). Todos aquellos hilos que no desarrollen la pista del hilo enfocado o alguna de sus ancestros serán considerados hilos secundarios.

A medida que evolucionan los turnos de los participantes, durante los procesos de interpretación y generación del diálogo, la pista de los turnos que desarrollan será actualizada en función de la relación de la pista del hilo que se encuentran desarrollando con respecto a la del hilo enfocado. De esta forma, la pista de un turno no es constante a lo largo de toda la contribución, pudiéndose pasar de desarrollar asuntos primarios a secundarios o viceversa (en función de los hilos que desarrolle el turno del propio participante, pero también de los que desarrolle el hablante).

5.2.2 Gestión de Metas

La gestión de metas es crucial para afrontar la interacción desde un desarrollo no basado en el *ciclo de la interacción* [Apartado 2.1]. Esta nueva función permitirá generar contribuciones del sistema cuando el estado de sus metas (tanto las propias como las combinadas) lo requiera (por variaciones en sus criticidades o compromisos, respectivamente). En consecuencia, la participación del sistema en la interacción no se limita a respuestas a las contribuciones previamente realizadas por el usuario, con lo que se abren las puertas al desarrollo de fenómenos como el solapamiento, la interrupción o la auto interrupción.

La gestión de metas incluirá la gestión de la inserción de metas discursivas propias del sistema, la monitorización del compromiso y criticidad de las metas a lo largo de la interacción y la gestión de su cancelación.

5.2.2.1 Monitorización de Hilos Combinados

Cuando el compromiso de alguno de los hilos es actualizado, tras procesos de interpretación (incluida la reinterpretación) o de confirmación, y según define la formalización del hilo en el dominio de interacción, el Monitor de Metas consultará el nivel de compromiso alcanzado en sus distintas componentes: atención, interés o información. Ante caídas críticas en alguno de dichos aspectos el Monitor de Metas insertará metas propias del sistema encaminadas a restablecer su nivel.

El sistema desarrollará pausas o actos nulos para reforzar la atención, el compromiso o el interés cuando su caída no es especialmente brusca. Para un refuerzo mayor, y cuando la caída afecta a la información o a la atención, puede desarrollar hilos de redundancia o aumentar el nivel de información que acompaña a los actos comunicativos de sus contribuciones. En esta misma línea, también podrá enumerar la línea de ancestros del hilo enfocado, lo que mejorará en especial la atención. Finalmente, cuando la caída de alguno de estos aspectos es brusca, el sistema recurrirá a interrupciones directas. Algunos ejemplos son los siguientes:

- *Refuerzo de atención*: "Me he perdido. ¿De qué estamos hablando?".
- *Refuerzo de interés*: "¿Quiere seguir programando el aviso?"
- *Refuerzo de información*: "No me refería a la hora de la reunión, sino a la de la comida".

5.2.2.2 Monitorización de Metas Discursivas Propias

Cada acción combinada en la interacción está compuesta por las acciones individuales de cada uno de los participantes. En el caso de las acciones correspondientes al sistema, éstas se encuentran representadas en el espacio de metas discursivas propias del Gestor de Metas.

Las metas discursivas propias del sistema surgen de la necesidad de alguno de los componentes del sistema de incluir una nueva meta en la interacción (que terminará desembocando en la interacción para ser comprometida o no por el resto de participantes), o bien del compromiso del sistema a desarrollar alguna de las metas propuestas por otros participantes (cuando sus metas son comprometidas). En lo que respecta a las metas discursivas propias del sistema, el Gestor de Metas podrá recibir metas desde los diversos componentes que integran el sistema de interacción a través del servicio de inserción de nuevas metas. Las solicitudes de inserción de metas provienen de los diversos componentes de la arquitectura, quienes, en determinados momentos, pueden requerir el desarrollo de metas particulares. Estas metas serán insertadas tanto para poder resolver solicitudes pendientes de otros componentes (por ejemplo, procesadores de lenguaje natural [Apartado 4.1.1.1]), como por iniciativa propia del sistema de interacción (Auto Modelo, Situación [apartados 4.2.3.5 y 4.2.3.3, respectivamente], etc.). Algunos ejemplos de metas que pueden insertar los distintos componentes del sistema son: solicitar información al usuario, cuando la aportada en la interacción hasta el momento es insuficiente ("*¿A qué hora quiere programar la alarma?*"); avisarle de eventos espacio-temporales ("*¡Cuidado!, el suelo está mojado*"), comunicarle problemas de interpretación "*hay mucho ruido, no le entiendo*"); etc.

Cada una de las metas discursivas propias del sistema será desarrollada a través de alguno de los hilos definidos en el dominio de interacción y está caracteriza por un determinado valor de criticidad que varía con el tiempo. La criticidad de una meta viene dada por una función definida en el momento de inserción de la meta por el componente que la insertó. Esta función describe la forma en la que la criticidad de la meta evoluciona a medida que el estado de interacción y el conocimiento sociolingüístico cambian. Cuando las distintas variables manejadas por los distintos modelos de conocimiento actualizan su valor (bien por efecto de contribuciones de otros participantes o de las del propio sistema, o por cambios en la situación, etc.), la criticidad de las metas discursivas propias del sistema aumentará o disminuirá. Al cambiar el valor de criticidad de una meta (bien al aumentar o al reducirse) el Gestor de Metas enviará una nueva solicitud de formalización al componente Coordinador de Procesos [Apartado 5.1.2]. Ante estas solicitudes, se desencadenará, en primer lugar, un proceso de decisión de toma de turno para calificar la urgencia con que debe ser desarrollada dicha meta (en el componente Gestor de Toma de Turno [Apartado 5.2]), y posteriormente un proceso de formalización de contribución (en el Generador de Diálogo [Apartado 5.2.4]). En función del estado de las metas, de los turnos, de la palabra y del resto de conocimiento sociolingüístico involucrado, el hilo del que depende la consecución de la meta podrá ser desarrollado o no a través de una nueva contribución de sistema (que a su vez puede consistir en una reformulación o rectificación). Cuando las variables pertenecen a modelos de conocimiento cuyo estado es versionable y es posible navegar entre ramas alternativas del árbol o retroceder a versiones anteriores (como es el caso del Estado de Interacción [Apartado 5.2.1]), también se recibirán notificaciones de cambio de dichas variables. Esto ocurrirá como consecuencia de cambios de versión producidos tras procesos como el de reinterpretación.

Si a lo largo de la interacción se alcanza la resolución de una de las metas discursivas propias del sistema, el componente que la solicitó (Generador de Diálogo, procesadores de lenguaje natural, Modelo de Situación, etc. [apartados 5.2.4, 4.1.1 y 4.2.3, respectivamente]) será informado. En dicha notificación se le dará a conocer si la resolución de la meta supone un éxito particular para los intereses del sistema (se siguió la secuencia preferida [116]) o no. Del mismo modo, el componente que insertó una meta puede requerir su cancelación. Por ejemplo, el Generador de Diálogo, quien puede insertar metas como consecuencia de haber alcanzado determinados estados de interacción, podría requerir su cancelación a causa de que ocurran procesos de reinterpretación que deshagan la versión que motivaba su apertura. En estos casos el Gestor de Metas la eliminará de su espacio de metas discursivas propias, aunque la gestión de la cancelación de una meta podría requerir la inserción de metas específicas de cancelación cuando la meta ya había sido comprometida por los participantes y se había comenzado a

desarrollar en la interacción ("*en ese caso no tengo nada que decir*", "*pues entonces nada*", etc.).

Para cada una de las interacciones desarrolladas por el sistema, éste mantiene un espacio de metas discursivas propias específico. Existe, además, un espacio de metas discursivas propias del sistema no asignadas a sesión. Éste es gestionado por el Auto Modelo [Apartado 4.2.3.5]. Así, habrá tantos espacios de metas discursivas propias de sistema como interacciones desarrolla simultáneamente, más el espacio de metas no asignadas a ninguna interacción. Al recibir nuevas metas, el Gestor de Metas las incluye en el espacio de metas individuales genérico, y sólo cuando la criticidad de dichas metas es suficientemente elevada como para requerir su desarrollo, el Gestor de Metas determinará a través de qué sesión será tratada (en función de la capacidad de los interlocutores para ayudar a resolver dicha meta). En el caso de que los interlocutores con los que se desea desarrollar una meta discursiva propia del sistema no se encuentren interactuando con el sistema, éste iniciará con ellos una nueva interacción por iniciativa propia ("*Disculpe que le interrumpa, pero tiene una cita programada.*"). En ocasiones, la interacción finaliza sin que todas las metas discursivas del sistema hayan podido ser resueltas. En ese caso, el Gestor de Metas reasignará las metas aún abiertas al espacio de metas no asignadas a sesión. En la medida en que su criticidad vuelva a dispararse, procederá, tal y como se describió, para asignarla a alguna de las sesiones posibles.

5.2.3 Intérprete de Diálogo

El Intérprete de Diálogo es el componente que ofrece al Gestor de Continuidad [Apartado 5.1.1] el servicio de interpretación de nuevos actos comunicativos, el cual es solicitado a través del Coordinador de Procesos [Apartado 5.1.2].

A medida que el Gestor de Continuidad procesa los nuevos fragmentos de contribución que recibe de los Componentes de Entrada de la Interfaz [Apartado 4.1.1.1] podrá ir alcanzando porciones de contribución interpretables a nivel de diálogo. Cuando esto ocurre, el Gestor de Continuidad representa dichas porciones de contribución en forma de actos comunicativos y solicita su interpretación de diálogo al Intérprete de Dialogo a través del Coordinador de Procesos. La interpretación diálogo de cada secuencia de actos comunicativos de la contribución completa será tratada de forma atómica, como una única sección crítica, en exclusividad de acceso sobre el estado de interacción y sobre el resto de conocimiento sociolingüístico involucrado en ella.

Durante la interpretación de diálogo se analizan los progresos que desencadena la nueva secuencia de actos comunicativos sobre el estado de interacción y el resto de conocimientos

implicados. Este proceso conlleva analizar, para cada uno de los actos comunicativos recibidos, si supone progresos para alguno de los hilos combinados; si supone la apertura, reapertura o cierre de alguno otro hilo; y la forma en la que actualiza el conocimiento estático de la interacción (especialmente su contexto) y el estado de los turnos.

El análisis de los progresos que supone un acto comunicativo sobre el estado de alguno de los hilos, denominado interpretación estructural, consiste en evaluar la medida en la que el acto comunicativo puede hacer progresar alguno de los hilos de la zona común. Para ello, los hilos serán analizados según su orden en la estructura focal, en busca de alguno sobre el que el acto comunicativo pueda, por si sólo o en combinación con otros actos comunicativos de la secuencia, suponer una transición a otro estado. Consiste en un estudio de los progresos en la interacción a nivel local.

Un acto comunicativo, por si sólo o junto a otros de los actos comunicativos de la secuencia, podría suponer la apertura, reapertura o cierre de hilos en la zona común y la actualización de la estructura intencional. Este análisis, denominado interpretación dinámica, permite actualizar la estructura global de la interacción, detectando el desarrollo de nuevos hilos, su cierre o recuperación por iniciativa del interlocutor. Se tratará del cierre de un hilo cuando se alcance alguno de sus estados finales (permitiendo la resolución de la meta que representa), será reapertura cuando haga progresar a un hilo que previamente se cerró y será apertura cuando se cumplan las condiciones (encontrarse en un estado de un determinado hilo que, dado el reparto de roles, permita ajustar al nuevo hilo como hijos suyo) bajo las cuales el acto comunicativo (por si solo o junto a otros de los actos comunicativos de la secuencia) se corresponde con alguna de las iniciativas contempladas para alguno de los hilos del dominio de interacción.

Los actos comunicativos incluyen, por otro lado, información estática que deberá registrarse en la interacción. Por ello, con cada acto comunicativo deberá actualizarse el contexto (el del hilo propio sobre el que será ajustado el acto comunicativo) y la historia de la sesión.

Durante el proceso de interpretación de diálogo se actualizará también el turno del interlocutor que está desarrollando la contribución en el Gestor de Toma de Turno [Apartado 5.1.3]. Esta actualización implica actualizar su estado (proyectar, estimar o detectar la expresión de posibles *TRP*), su *pista de acción* (primaria o secundaria, en función de la relación entre la pista de los hilos desarrollados y el hilo enfocado), y si supone la solicitud de turno o la designación de un hablante siguiente.

Además, previamente a la propia interpretación de cada uno de los actos comunicativos, deberán ser resueltas las posibles referencias deícticas que pudieran contener los actos comunicativos. La resolución de referencias deícticas será realizada apoyándose en componentes como el de Sesión (para la resolución de deíxis de discurso, anáforas, catáforas, etc.) y el de Situación (para la deíxis temporal, de objetos, social, etc.) [apartados 4.2.3.2 y 4.2.3.3, respectivamente].

5.2.3.1 Reinterpretaciones

El no disponer de contribuciones completas durante la interpretación de diálogo de cada una de las secuencias recibidas hace necesario contemplar el tratamiento de las reinterpretaciones. Se denominará reinterpretación a la revisión de interpretaciones de diálogo realizadas previamente como consecuencia de la obtención de actos comunicativos más precisos que los que previamente fueron interpretados (refinados a medida que van siendo adquiridos nuevos fragmentos de la contribución del interlocutor) ,o por la elección de una mejor línea de interpretación de entre todas las posibles (al hacerse más probables al recibirse nuevos actos comunicativos).

El Gestor de Continuidad produce actos comunicativos de forma incremental. Esto significa que, a medida que la propia contribución del interlocutor avanza, obtendrá iterativamente estimaciones más precisas de las acciones comunicativas que se esconden tras la contribución del interlocutor. Las nuevas versiones obtenidas de los actos comunicativos pueden modificar la propia acción comunicativa por completo, o afectar exclusivamente a los parámetros lingüísticos que lo acompañan.

Por otro lado, debe ser considerado que la *interacción natural* no es una ciencia exacta, por lo que los actos comunicativos podrían admitir interpretaciones de diálogo diversas sobre el mismo estado de interacción de partida. En muchos casos, una misma acción comunicativa puede ser interpretada, a la vez, como el progreso, apertura, reapertura o cierre de varios hilos a la vez. En estos casos, estimar la verdadera intención del interlocutor se hace complicado, y aunque puede ser realizado un cálculo de probabilidades preeliminar sobre las distintas alternativas, a medida que la contribución progresa y se reciben nuevos actos comunicativos, alternativas poco probables pueden convertirse en la línea de interpretación que permite un mejor encaje de los actos comunicativos. Esto obliga al Intérprete de Diálogo a mantener interpretaciones paralelas de las contribuciones que se están cursando, a elegir de entre todas ellas la más probable en cada momento y a revisar las interpretaciones alternativas cada vez que

es analizado un nuevo acto comunicativo, con objeto de detectar interpretaciones incorrectas de actos comunicativos previos.

En ambos casos, la reinterpretación de diálogo será soportada por la gestión de versiones del Estado de Interacción [Apartado 5.2.1]. En el primero de los casos se desharán los progresos alcanzados como consecuencia de la interpretación de los actos comunicativos obsoletos y se llevará a cabo la interpretación de la nueva versión de los actos comunicativos producidos por el Gestor de Continuidad sobre la última versión previa al acto comunicativo confirmado. Para el segundo de los casos, el intérprete se apoya en la capacidad del Estado de Interacción [Apartado 5.2.1] de soportar líneas de interpretación alternativa, pudiendo descartar la línea de progreso más probable hasta en cada momento en beneficio de otra en la que los nuevos actos comunicativos "tengan un mayor sentido" (es decir, puedan ser encajados en mejores condiciones sobre el estado de interacción).

La habilidad para deshacer progresos alcanzados por actos comunicativos obsoletos, o para desplazar la versión actual de una rama del árbol de versiones del Estado de Interacción a otra, puede hacer inconsistentes los progresos alcanzados por procesos desarrollados posteriormente. El retroceso a versiones anteriores del estado de interacción, o el cambio de rama del árbol de versiones, pueden provocar la desaparición de los progresos a partir de los cuales se desarrollaban la apertura, reapertura o cierre de determinados hilos. También su desarrollo durante los procesos de interpretación de diálogo. Del mismo modo, determinadas metas que fueron introducidas en la interacción por el propio sistema como consecuencia de estados descartados pueden quedar ahora injustificadas.

Por ello se requiere recuperar y atender de nuevo aquellas solicitudes posteriores al estado descartado (aun a pesar de desarrollarse con retraso). Se volverá a reinterpretar y confirmar todo lo que ocurrió con posterioridad a la interpretación revisada, detectándose las divergencias en los estados finales alcanzados. Con ello se hace posible identificar interpretaciones que, dado el nuevo estado de interacción, sean erróneas. Del mismo modo, se comunicará al Generador de Diálogo [Apartado 5.2.4] qué hilos han dejado de tener sentido y qué progresos se han deshecho, permitiéndole cancelar aquellas metas discursivas que introdujo y que han dejado de tener sentido. Si las contribuciones que realizó por el error de interpretación pudieran estar fuera de lugar o haber sido malinterpretadas, el sistema podrá desarrollar las aclaraciones, disculpas o rectificaciones oportunas a través de la inserción de nuevas metas discursivas del sistema en el Gestor de Metas [Apartado 5.2.2].

5.2.4 Generador de Diálogo

La propuesta requiere un componente Generador de Diálogo capaz de desarrollar procesos de generación, no como consecuencia de una fase previa de interpretación de diálogo, sino por evolución de las metas internas del sistema; de los hilos comprometidos; y de los turnos de la interacción a unos estados que hacen relevante una toma de turno en la interacción por parte del sistema. Se requiere también la capacidad de desarrollar procesos de generación simultáneamente a otros posibles procesos que pudieran estar ocurriendo en la *interacción natural* (actualización de conocimiento circunstancial, interpretación de las contribuciones de otros participantes, etc.). Del mismo modo, debe ser contemplada la posibilidad de reformular, interrumpir o auto interrupir la contribución del propio sistema mientras ésta se está siendo producida.

Todo ello requiere descomponer la generación de diálogo en dos subprocesos: *formulación de contribuciones del sistema* y *confirmación de la síntesis de las expresiones que las componen*. La formulación de contribuciones se desencadena ante solicitudes remitidas por el Coordinador de Procesos [Apartado 5.1.2]. Comprende una decisión de toma de turno (realizada por el Gestor de Toma de Turno [Apartado 5.1.3]) y el diseño de la contribución con la que el sistema pretende hacer progresar la interacción (de lo que se encarga el Generador de Diálogo). Por su parte, la confirmación de síntesis conlleva ejecutar los progresos de la contribución en el estado de interacción, pero sólo en la medida en la que los Componentes de Interfaz de Salida [Apartado 4.1.1.2] van confirmando que las porciones de contribución que constituyen dichos progresos (aquellas que representan actos comunicativos completos) han sido expresadas.

5.2.4.1 Formalización de contribuciones

Tras una decisión de toma de turno, será competencia del Gestor de Diálogo el formalizar una nueva contribución del sistema. Durante el proceso de decisión de toma de turno, el componente Gestor de Toma de Turno [Apartado 5.1.3.4] marca cada una de las distintas metas de la interacción con un determinado nivel de urgencia (de acuerdo al estado de los turnos, el de posesión de la palabra, los candidatos a tomarla, el estado de las metas de la interacción y el resto de conocimiento sociolingüístico implicado). Tras esto, es competencia del Generador de Diálogo determinar los movimientos que debe realizar el sistema en la interacción y construir la nueva contribución del sistema correspondiente a dichos movimientos. El resultado podrá ser una nueva contribución, la reformulación o auto interrupción de la contribución en curso, o un turno de paso.

Determinar el progreso de un hilo consiste en construir las versiones del estado de interacción y resto de conocimiento sociolingüístico que surjan de la ejecución de las tareas asociadas al rol del sistema en cada uno de los estados por los que transcurre (haciendo uso de las funciones del Gestor de Tareas [Apartado 4.2.2]) y en añadir al discurso de la contribución correspondiente al hilo los actos comunicativos que vayan desprendiéndose de sus transiciones. Estas serán las versiones que posteriormente serán confirmadas al recibirse la confirmación de expresión de cada uno de los actos comunicativos de la contribución del sistema.

Existen situaciones en las que el desarrollo de un hilo de urgencia elevada requiere el desarrollo previo de alguno de sus hijos para progresar, o incluso el desarrollo de sus ancestros. En estos casos, dichos hilos serán desarrollados incluso sin ser urgentes. Un ejemplo es el desarrollo de la propia conversación (el hilo base), cuya urgencia puede dispararse ante silencios prolongados. Para hacerla progresar, el sistema tratará de desarrollar alguno de los asuntos que todavía no se hubiesen cerrado (es decir, alguno de sus hilos hijos aun abiertos).

En determinados casos, el progreso de un hilo que hubiese alcanzado una criticidad elevada no depende del sistema (para cuyo rol pueden no estar descritas transiciones desde el estado en el que se encuentra). En dichos casos el Generador de Diálogo incluirá nuevas metas discursivas propias en el Gestor de Metas con el objetivo de favorecer participaciones del los interlocutores en la línea de hacerlo progresar ("*entonces, ¿qué opinas de…?*"). En el caso del hilo base, suelen introducirse metas de relleno (como hablar del tiempo).

Posteriormente, los actos comunicativos producidos por el sistema son sometidos a un proceso de generación de referencias deícticas. Para ello se aplicarán las capacidades del resto de modelos, como el de sesión o el de situación, con lo que será posible dotar a la contribución formalizada por el sistema de una mayor naturalidad.

Todas estas consideraciones darán lugar a una nueva contribución del sistema que será enviada al Gestor de Continuidad [Apartado 5.1.1], quién gestionará la continuidad de su expresión. Tras ello el Adaptador Multimodal [Apartado 4.1.3.1] diseñará su expresión en modalidades y lenguaje naturales a sus interlocutores. Finalmente, el resultado será encaminado hacia los Componentes de Interfaz de Salida [Apartado 4.1.1.2], para ser expresado a lo largo del desarrollo de un turno.

Si durante el desarrollo del turno se producen cambios que pudieran afectar a la urgencia de las metas; al estado de los turnos; a la posesión de la palabra; o a los candidatos a tomarla, la expresión de dicho turno será inmediatamente suspendida y se desarrollará una nueva decisión de toma de turno (y una posterior formalización), que podrá resultar en: la

continuación del turno (si la contribución resultante es igual a la parte de contribución que aun no había sido expresada); una reformulación (si la contribución resultante conecta con la parte ya expresada pero es distinta a la que quedaba por expresar); una rectificación (si modifica lo ya expresado); o una auto interrupción (si supone una ruptura con lo ya expresado o una cancelación del turno).

5.2.4.2 Confirmación de Síntesis

Los progresos producidos sobre la interacción por la contribución formulada por el sistema sólo serán llevados a cabo una vez que puedan darse por expresados los fragmentos de contribución que los constituyen. Al considerarse el acto comunicativo la mínima unidad de discurso que puede producir progresos en la interacción, cada vez que el Gestor de Continuidad, junto con el Adaptador Multimodal y procesadores de lenguaje natural [apartados 5.1.1, 4.1.3.1 y 4.1.1.2, respectivamente], confirman la expresión de uno de los actos comunicativos de la contribución del sistema, desencadena una solicitud de confirmación de expresión de acto comunicativo al Generador de Diálogo, que será tramitada a través del Coordinador de Procesos [Apartado 5.1.2].

La confirmación de expresión de un acto comunicativo ejecutará sobre el estado de interacción y sobre el resto de conocimiento sociolingüístico las actualizaciones asociadas a ese acto comunicativo. La confirmación de expresión de un acto comunicativo consiste en confirmar la versión del estado de interacción, así como del resto de modelos de conocimiento, que fue previamente diseñada durante la formalización y asociada al acto comunicativo que se expresó [apartados 5.2.1 y 4.2.3, respectivamente]. La confirmación podrá conllevar la actualización de la estructura focal, de la estructura intencional, del estado de desarrollo de un hilo y la actualización del contexto, entre otros.

A medida que son confirmados nuevos actos comunicativos, el Generador de Diálogo deberá también actualizar el estado del turno del sistema en el componente Gestor de Toma de Turnos. Deberá indicarse si, tras la confirmación de un acto comunicativo, la actividad del turno continúa, si se ha generado una *TRP* o si se ha alcanzado lo que otros participantes podrían identificar como su proyección. Del mismo modo, con cada confirmación deberá actualizarse la pista desarrollada por la contribución del sistema y la posición de los hablantes frente a turnos futuros, en función de que el acto comunicativo confirmado contenga designaciones de hablantes siguientes (bien de forma directa, bien de forma indirecta) o solicitudes de turno del sistema. La actualización de estos parámetros en el Gestor de Metas [Apartado 5.2.2] puede desencadenar la actualización en el estado de posesión de la palabra. En la medida en que esto

suceda se tramitarán nuevas solicitudes de formulación con el objetivo de adaptar la contribución del sistema a la nueva situación.

5.3 *DESCRIPCIÓN DE PROCESOS SEGÚN LA PROPUESTA*

A medida que se van desarrollando los procesos de interpretación de actos comunicativos de los interlocutores, o de generación de los del sistema, se producen cambios sobre el estado de interacción; el estado de las metas y la toma de turno; y sobre el resto de conocimiento sociolingüístico de influencia en la interacción [Apartados 4.2.1.1, 5.2.2, 5.1.3 y 4.2.3, respectivamente]. Estos cambios también ocurren como consecuencia de la inserción o cancelación de metas discursivas propias del sistema, que pueden ser realizadas por cualquiera de los componentes de la arquitectura. Todos estos fenómenos determinan la forma en que se desarrolla la toma de turno en la interacción [Figura 14].

Ante cada uno de estos cambios, el Gestor de Metas revisará el estado de las metas discursivas propias del sistema y el de las metas combinadas [Apartados 5.2.2.15.2.2.2]. Cuando, como consecuencia de dichos cambios, se producen variaciones significativas en la urgencia de las metas de la interacción, el componente Gestor de Metas desencadena nuevos procesos de generación. Esto mismo ocurrirá ante los cambios en el estado de los turnos, la posesión de la palabra, o los candidatos a tomarla. Los procesos de generación comienzan por una decisión de toma de turno, realizada por el Gestor de Toma de Turnos [Apartado 5.1.3], que determinan qué hilos están en condiciones de ser desarrollados en la interacción y bajo qué nivel de urgencia (dado el estado de los turnos, el estado de posesión de la palabra, quienes son los candidatos a tomarla y el resto de conocimiento sociolingüístico de influencia en la interacción). Tras una decisión de toma de turno, el Generador de Diálogo [Apartado 5.2.4] revisará los progresos que debe realizar el sistema en la interacción a los niveles estático, dinámico y estructural, obteniendo como producto una nueva contribución de sistema. En caso de que el sistema no formalice una nueva contribución, pero esté en posesión de la palabra, generará una contribución de turno de paso.

La contribución formalizada es remitida al Gestor de Continuidad [Apartado 5.1.1], para que la combine de la forma más fluida posible con la contribución previa que podría estar generando el sistema. El Gestor de Continuidad también diseñará la expresión de aquellos marcadores de toma de turno que se produzcan como alteraciones en la continuidad de la contribución. Tras ello, el Adaptador Multimodal [Apartado 4.1.3.1] define las modalidades y la

lengua en la que debe ser expresada la contribución, y el resultado es remitido a los Componentes de Interfaz de Salida [Apartado 4.1.1.2].

Los Componentes de Interfaz de Salida sintetizan, a lo largo de un turno del sistema, la contribución formalizada. A medida que pueden darse por expresados cada uno de los fragmentos de expresión que la componen, las confirmaciones de síntesis se remiten al Generador de Diálogo, quién confirmará en el Estado de Interacción [Apartado 5.2.1] las versiones que van siendo alcanzadas con la expresión de cada uno de los actos comunicativos de la contribución. Simultáneamente, el Gestor de Toma de Turno va actualizando el estado del turno del sistema, el estado de posesión de la palabra y los candidatos a tomarla.

Durante la confirmación de expresión de una contribución formalizada, podrían ocurrir cambios en el estado de las metas (propias del sistema o las combinadas) o en el estado de la toma de turno, lo que puede suponer la pérdida de vigencia de la contribución en curso del sistema. En estos casos, la generación de la porción de contribución pendiente será automáticamente suspendida. Tras ello, se analizará bajo un nuevo proceso de decisión de toma de turno cuál es la actual urgencia de las metas de la interacción, y, con ello, el Generador de Diálogo procederá a formalizar una nueva contribución de sistema. En función de los cambios producidos el resultado podrá ser la continuación de la contribución en curso, su reformulación suave, su rectificación o una auto interrupción.

Simultáneamente a los procesos de generación, pueden darse procesos de interpretación de contribuciones de otros participantes en la interacción. Los procesos de interpretación parten de la adquisición, por los Componentes de Interfaz de Entrada [Apartado 4.1.1.1], de los fragmentos de expresión que los interlocutores van produciendo a lo largo de su contribución. Estos fragmentos, tras ser unificados en un flujo de datos único en el Adaptador Multimodal, se remiten al Gestor de Continuidad para ser combinados con los fragmentos anteriormente recibidos. El Gestor de Continuidad actualiza el estado del turno del interlocutor; el estado de posesión de la palabra; y los candidatos a tomar la palabra. También detecta marcadores de toma de turno expresados como alteraciones en la continuidad temporal del turno. Cuando el Gestor de Continuidad detecta fragmentos de contribución con sentido interactivo completo (actos comunicativos), desencadena procesos de interpretación de diálogo. El Intérprete de Diálogo [Apartado 5.2.3] analiza los progresos estáticos, estructurales y dinámicos que suponen los actos comunicativos recibidos en el estado de interacción y resto de conocimiento sociolingüístico asociado. La recepción de nuevos fragmentos de contribución puede desencadenar la reinterpretación de actos comunicativos previamente interpretados. En la medida en la que la reinterpretación afecte a progresos posteriores realizados sobre la

interacción, podría ser necesario realizar rectificaciones y cancelación de metas introducidas erróneamente en la interacción.

Finalmente, el Coordinador de Procesos [Apartado 5.1.2] se encarga de que los procesos de interpretación de nuevos actos comunicativos; la formalización de nuevas contribuciones (que comprende la decisión de toma de turno y la construcción de una nueva contribución); y las confirmaciones de síntesis, sean producidos de forma ordenada y bajo un control de acceso a los recursos compartidos.

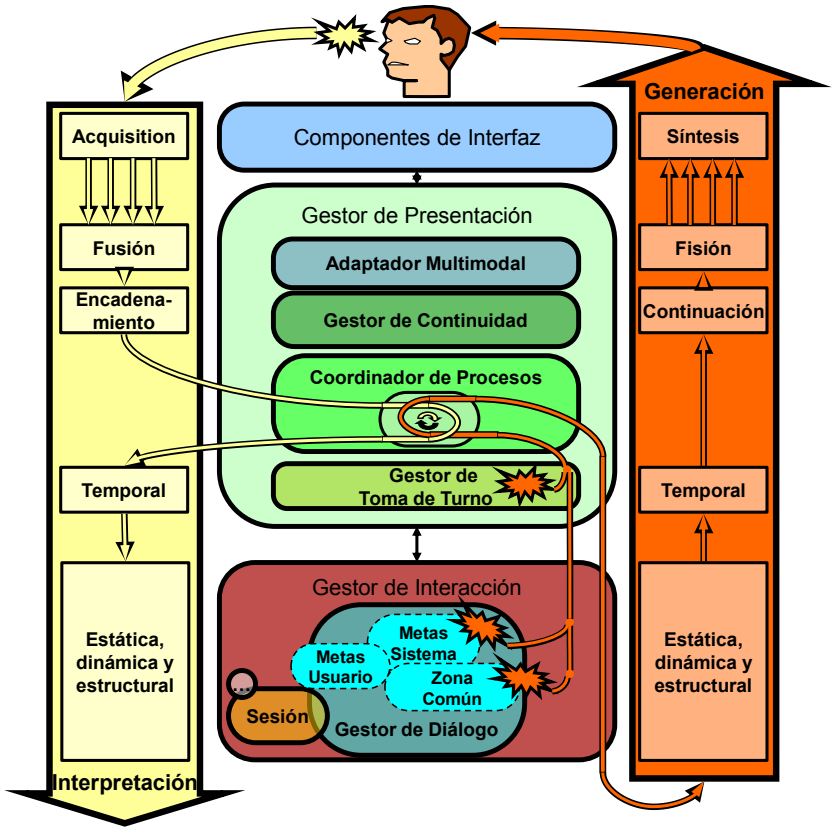


Figura 14: Comunicación entre componentes de la arquitectura en una toma de turno avanzada

Capítulo 6 DESARROLLO DE LA PROPUESTA

A lo largo del presente capítulo se profundizará en la forma en la que se aborda la funcionalidad de cada uno de los modelos de conocimiento que incluye la propuesta. Se incluyen tanto los nuevos componentes (Gestor de Continuidad [Apartado 6.1], Coordinador de Procesos [Apartado 6.2] y Gestor de Toma de Turnos [Apartado 6.3]), como las extensiones de los componentes Estado de Interacción [Apartado 6.4], Intérprete de Diálogo [Apartado 6.5] y Generador de Diálogo [Apartado 6.6]. Este apartado concluirá con la descripción de las soluciones propuestas para el componente Gestor de Metas [Apartado 6.7].

6.1 *GESTOR DE CONTINUIDAD*

El Gestor de Continuidad es el componente que se encuentra entre los Componentes de Interfaz [Apartado 4.1.1] y el de Adaptación Multimodal [Apartado 4.1.3.1]. Entre estos tres componentes se producen dos flujos distintos de comunicación. Uno de ellos el producido como consecuencia de la interpretación de las contribuciones de los interlocutores del sistema (que parte de los Componentes de Interfaz y se dirige hacia el componente de Adaptación Multimodal, pasando por el Gestor de Continuidad). El otro el que viene motivado por las contribuciones generadas por el propio sistema (recibido desde el Adaptador Multimodal y que se dirige hacia los Componentes de Interfaz, pasando por el Gestor de Continuidad). De esta forma, las funciones que aborda el Gestor de Continuidad se dividen entre las que afectan a la gestión de la continuidad de los procesos de interpretación de las contribuciones de entrada (las que producen otros participantes), y las que afectan a los de generación de las suyas propias. Cada uno de estos grupos de funciones será descrito a continuación.

6.1.1 Gestión de Continuidad en las Contribuciones de Entrada

Los Componentes de Interfaz de Entrada remiten al Gestor de Continuidad los nuevos fragmentos de contribución que reciben de sus interlocutores a través de alguna de las modalidades disponibles. El Gestor de Continuidad mantiene, para cada interlocutor y cada modalidad disponible, una memoria con los fragmentos de contribución anteriormente recibidos. Ante la recepción de nuevos fragmentos de contribución, el Gestor de Continuidad (en colaboración con el Adaptador Multimodal y los procesadores de lenguaje natural) detecta, en la porción de contribución que el sistema ha recibido hasta el momento, aquellas secuencias de fragmentos que, por sí mismas, tienen significado interactivo completo (*actos comunicativos*). Cuando esto ocurre, el Gestor de Continuidad solicita la interpretación de dichos actos comunicativos al Intérprete de Diálogo a través de Coordinador de Procesos. Durante este proceso, el Gestor de Continuidad se encarga también de la detección del comienzo y cese de la actividad de los interlocutores (información con la que actualiza el Gestor de Toma de Turnos) y de la detección de determinados marcadores de toma de turno [Figura 15].

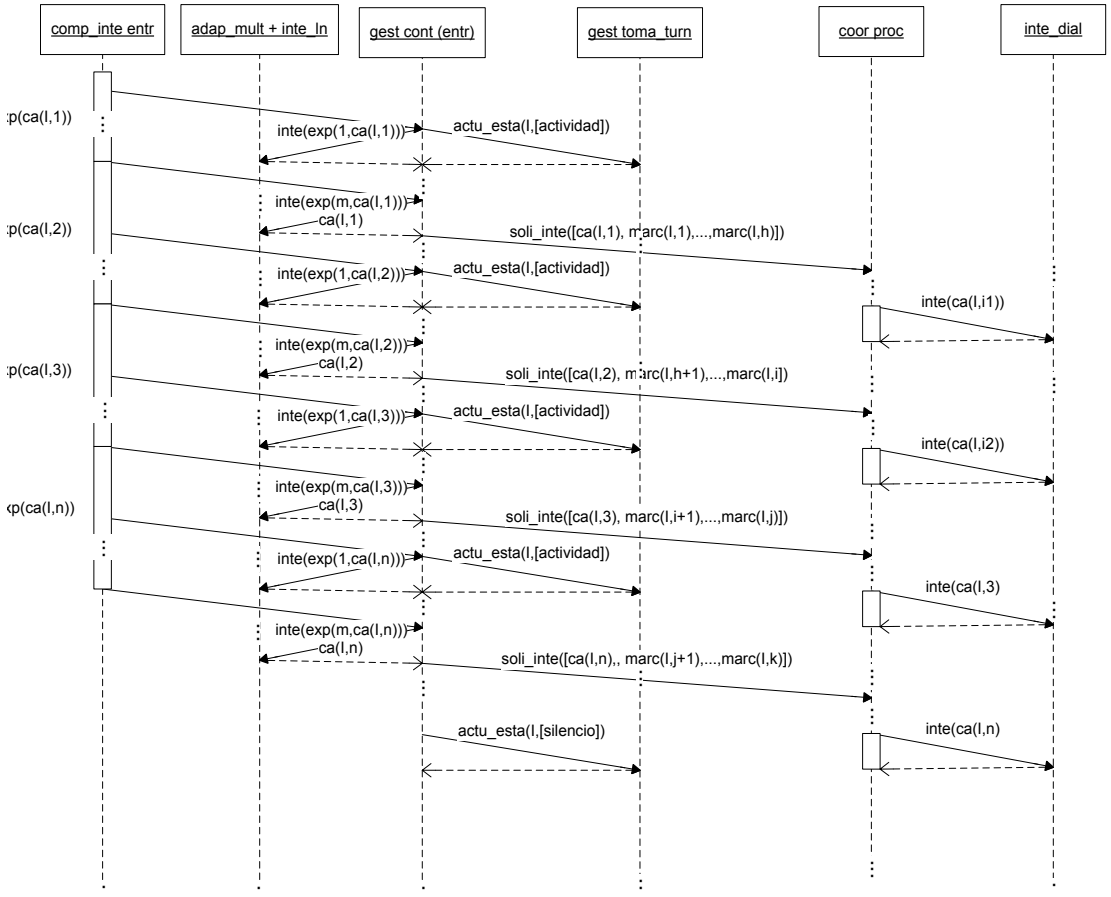


Figura 15: Diagrama de secuencia de gestión de continuidad de las contribuciones de entrada

6.1.1.1 Interpretación Incremental

A medida que se reciben fragmentos de contribución a través de la interfaz, el Gestor de Coordinación de Entrada combina estos fragmentos con los anteriormente recibidos para generar unidades mayores que puedan representar acciones comunicativas completas por si solas. Para ello, a su recepción, los fragmentos serán clasificados por modalidades y concatenados con los fragmentos de contribución más recientemente recibidos en la misma modalidad. Así, lo que de otro modo sería tratado como una unidad atómica (por ejemplo la posible intervención monomodal de usuario "*lístame qué avisos tengo pendientes para mañana*"), en una interpretación incremental sería recibida como una sucesión de fragmentos a lo largo del tiempo. Suponiendo, por sencillez, una granularidad de palabra, la contribución de este ejemplo sería recibida como la sucesión de fragmentos: "*lístame*", "*qué*", "*avisos*", etc.

Para cada fragmento recibido, el Gestor de Continuidad propone al Adaptador Multimodal (responsable de la fusión multimodal y coordinador de la interpretación de lenguaje natural) posibles combinaciones de estos fragmentos con los anteriormente recibidos. En realidad, las unidades que remite para su adaptación e interpretación de lenguaje natural serán la concatenación de este último fragmento con los n anteriormente recibidos (tomando n valores comprendidos entre 0 y un valor máximo dependiente de las limitaciones físicas del sistema). En el caso del ejemplo, en el momento en el que se recibe el fragmento "*para*", serían propuestas como unidades para su adaptación multimodal e interpretación de lenguaje natural las unidades "*para*" (sólo el último fragmento), "*tengo para*" (último fragmento con el anterior), y así hasta "*lístame qué avisos tengo para*" (con todos los fragmentos de la contribución). Esto mismo tratamiento se aplicaría también a las contribuciones multimodales.

El Adaptador Multimodal devolverá, en caso de existir, las posibles interpretaciones de lenguaje natural de cada una de las unidades propuestas. Los resultados de estas interpretaciones tienen forma de secuencias de actos comunicativos y van acompañadas de información adicional sobre la calidad de la interpretación realizada. En la medida de la calidad se consideran parámetros como la proporción de la unidad que no pudo ser encajada; los fragmentos de contribución que no pudieron ser encajados; y los fragmentos de contribución pendientes de interpretación [Ejemplo 15].

En este caso, la mejor interpretación posible al ser recibido el fragmento "*para*" sería <*El usuario solicita un listado de los avisos pendientes*>. Estos es así puesto que, de todos los resultados de interpretación de lenguaje natural propuestos, es el que da mayor tasa de encaje (4 palabras de 5). De esta forma, la secuencia de fragmentos de contribución "*Lístame qué avisos*

tengo", sería considerada una contribución de usuario con sentido completo (al menos provisionalmente) y se procedería su interpretación de diálogo (En el Intérprete de Diálogo, a través del Coordinador de Procesos). Quedará pendiente la interpretación del fragmento "*para*" hasta ser recibidos nuevos fragmentos.

Fragmentos descartados	Secuencia interpretada	Interpretación de lenguaje natural	Fragmentos no encajados en la interpretación	Tasa de encaje (encajadas / totales)
Lístame qué avisos tengo…	*…para*	--	--	0/5
Lístame qué avisos…	*…tengo para*	El interlocutor informa de que posee algo	*Para*	1/5
Lístame qué…	*…avisos tengo para*	El interlocutor informa de que conoce avisos	*Para*	2/5
Lístame…	*…qué avisos tengo para*	El Interlocutor solicita un listado de los avisos pendientes	*Para*	3/5
	Lístame qué avisos tengo para	El Interlocutor solicita un listado de los avisos pendientes	*Para*	4/5

Ejemplo 15: Resultados de la interpretación de lenguaje natural de la contribución "*lístame que avisos tengo*"

Con este enfoque se hace posible obtener interpretaciones parciales de una contribución e irlas refinando y corrigiendo a lo largo del tiempo (según vayan siendo recibidos nuevos fragmentos de contribución). Así, al ser recibido el fragmento "*mañana*", en el caso del ejemplo, podría obtenerse la nueva interpretación de la contribución en curso <*El Interlocutor solicita un listado de los avisos pendientes para mañana*>, de mayor precisión y tasa de encaje que la interpretación previa <*El usuario solicita un listado de* (todos) *los avisos pendientes*> (que ahora supone descartar los fragmentos "*para*" y "*mañana*"). Ante esta nueva interpretación de lenguaje natural, el Gestor de Continuidad solicitará al Intérprete de Diálogo, a través del Coordinador de Procesos, una nueva interpretación de diálogo que rectificará la interpretación previamente realizada.

Junto a esto, en el uso espontáneo del lenguaje son frecuentes la ocurrencia de discontinuidades ("*Lístame qué avisos tengo para…. para… para mañana*"), rectificaciones ("*Lístame qué avisos tengo para mañana… para pasado*") y otros tipos de alteraciones de la continuidad de la contribución [12]. Es también función del Gestor de Continuidad eliminar dichos fenómenos de las secuencias de fragmentos que envía al Adaptador Multimodal y los procesadores de lenguaje natural para facilitar su interpretación de lenguaje natural.

6.1.1.2 Detección de Actividad e Inactividad

El significado que tiene un conjunto de fragmentos de contribución, provisional o no, sólo es obtenido posteriormente a su interpretación de diálogo. De ninguna forma antes. Sólo cuando se han recibido suficientes fragmentos de la contribución como para considerarla con sentido interactivo completo es posible realizar con ella una interpretación a nivel de diálogo (con las nuevas secuencias de actos comunicativos obtenidas). Del ejemplo "*lístame los avisos que tengo para mañana*", sólo tras alcanzarse "*lístame los avisos que tengo*" se obtenía una primera interpretación de lenguaje natural. Hasta entonces, el Gestor de Continuidad no habría notificado que el turno del interlocutor se encuentra en un estado de actividad, por lo que, durante varios segundos, el sistema consideraría que la palabra sigue vacante y podría tratar de tomarla por error, provocando interrupciones o solapamientos no justificados. En definitiva, sin considerar el estado de actividad, se pierde información de gran importancia para la correcta representación del estado de los turnos, la posesión de la palabra y los candidatos a tomarla.

Para solventar este problema, el Gestor de Continuidad notifica al Gestor de Toma de Turno aquellas situaciones en las que existen fragmentos de contribución que, aun habiendo sido recibidos, no han producido, por el momento, resultados a nivel de interpretación de lenguaje natural. Para gestionar en qué momentos se envían notificaciones de actividad al Gestor de Toma de Turno, el Gestor de Continuidad asigna una máquina de estados [Figura 16] a cada uno de los participantes de la interacción. En lo que respecta a las máquinas de estado de los interlocutores del sistema, con cada nuevo fragmento se considera si con él pudieron ser obtenidos nuevos actos comunicativos; si en ese caso quedaron fragmentos de contribución pendientes de ser interpretados; y si existe una notificación de actividad previa (posterior al último acto comunicativo que fue interpretado). De analizar todos los posibles casos [Tabla 7], las situaciones en las que el Gestor de Continuidad debe notificar la actividad en un turno son aquellas en las que existen fragmentos de contribución que aun no pudieron ser interpretados y en que no existe ninguna notificación posterior a la obtención del último acto comunicativo.

De igual importancia a la detección de la actividad es la detección de la inactividad en el turno de los interlocutores. Los *silencios* pueden ser utilizados para marcar el final de una contribución y, a menudo, se emplean como recurso lingüístico. Desde el punto de vista de la toma de turno, puede ser utilizado, por ejemplo, para reclamar la atención (el profesor se calla para llamar la atención de los alumnos), para separar ideas (*lapsos*) o para liberar la palabra (*intervalos*). Por ello, tan importante como la notificación de la actividad de los interlocutores es la notificación de sus silencios. Con este fin, el Gestor de Continuidad monitorizará el tiempo transcurrido desde la recepción del último fragmento de contribución de un determinado

interlocutor. A este respecto, algunos estudios experimentales [92; 15] sitúan la máxima duración de estos silencios en aproximadamente un segundo, aunque este tiempo es muy dependiente de cuestiones culturales, circunstanciales y de la propia caracterización del interlocutor.

Tabla 7: Posibles casos de notificación de la actividad de un turno

Actividad notificada	AC	Fragmentos Pendientes	Notificar actividad
No	No	No	Sí
No	No	Sí	Sí
No	Sí	No	No
No	Sí	Sí	Sí
Sí	No	No	No
Sí	No	Sí	No
Sí	Sí	No	No
Sí	Sí	Sí	Sí

→

Actividad notificada	AC	Fragmentos Pendientes	Notificar actividad
No	No	X	Sí
X	Sí	No	No
X	Sí	Sí	Sí
Sí	No	X	No

Dado que determinados casos de titubeo o de repetición de palabras pueden denotar problemas en la formalización de la contribución (o un intento por retener injustificadamente la palabra), también será relevante su notificación al Gestor de Toma de Turno. Estas situaciones son denotadas *silencios oralizados* y es el Gestor de Continuidad el responsable de su detección. Serán detectados silencios oralizados ante la ocurrencia de repeticiones en las últimas palabras de la contribución ("*avisame a las… a las…*") o cuando aparezcan expresiones de relleno ("*avisame a las… eh… esto…*").

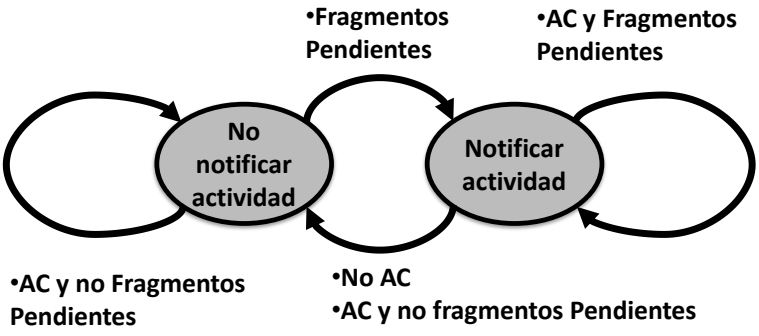


Figura 16: Máquina de estados de notificación de actividad

Los silencios y los silencios oralizados, al igual que la actividad, son notificados al Gestor de Toma de Turno para actualizar el estado del turno del participante que lo produjo.

6.1.1.3 Marcadores de Gestión de Toma de Turno

Entre los marcadores que los participantes pueden introducir en sus contribuciones están los relacionados con la gestión de la toma de turno. Aunque muchos de ellos se obtienen directamente de la interpretación de diálogo de la contribución (y por tanto su detección será realizada en el Intérprete de Diálogo), otros consisten en un conjunto de alteraciones sobre la continuidad temporal de la contribución que deben ser detectados por el Gestor de Continuidad. Entre ellos se encuentran:

- Solicitudes de palabra realizadas por *retardos momentáneos del turno* o por *reinicios del turno*:
    - Retardo momentáneo del turno: El hablante puede requerir la atención del oyente retrasando momentáneamente parte de la presentación. Esta técnica [Ejemplo 16] se suele utilizar cuando el interlocutor no mira al hablante [71].
    - *Lee: Puedes alcanzarme (0.2)*
      *Ray: *comienza a prestar atención**
      Lee: ese cartón de leche

      Ejemplo 16: Retardo momentáneo del turno

    - Reinicio del turno: Cuando el hablante empieza a hablar [Ejemplo 17] comprueba que el interlocutor le está mirando. Si no es así puede conseguir su atención, por ejemplo, reiniciando el turno [71].
    - *Lee: Puedes alcanzarme (0.2)*
      *Ray: *comienza a prestar atención**
      *Lee: *Puedes alcanzarme* ese cartón de leche*

      Ejemplo 17: Reinicio del turno

    - Existen otros marcadores que, por no estar relacionados con el Gestor de Continuidad, no serán tratados. Entre ellos se encuentran, por ejemplo, levantar la mano y expresar determinadas frases protocolarias ("*disculpe*"). La competencia de su reconocimiento recaería sobre los procesadores de lenguaje natural y el Adaptador Multimodal.
- Retenciones de la palabra:

  - Parada y Continuación: Cuando el hablante no sabe como continuar o aún no ha formulado el resto del desarrollo de su turno. En estos casos puede introducir silencios para evitar perder la palabra [26; 27; 28].
  - Comenzar y reparar: Los hablantes, a menudo, cambian de parecer acerca de lo que están presentando [115] por revisión del mensaje, reformulación de la contribución o detección de errores (léxicos, sintácticos, semánticos o en la articulación de la contribución). En estos casos es frecuente repetir desde alguna de las últimas palabras anteriores a la porción de contribución a ser reparada ("*Lístame los avisos para mañana… para pasado*" o "*Quiero poner la alarma a las 8… a las 7*").
  - También puede llevarse a cabo aplicando frases hechas o realizando señales específicas, como movimientos de manos (mostrar las palmas de las manos a los demás participantes pareciendo querer decir "*espera*") [44].
- Cesión de palabra:
  - Silencios: En muchos casos, los silencios se producen con el objetivo de denotar el final de una contribución en curso.
  - Junto a éste, existen otros marcadores que, sin ser competencia de Gestor de Continuidad, también conviene mencionar. Su reconocimiento está asociado a los niveles de adaptación multimodal y procesamiento de lenguaje natural. Son comenzar a prestar atención a algún otro participante; marcar cierta entonación; arrastrar la sílaba tónica o final; determinados movimientos de manos; o frases hechas y expresiones fáticas estereotipadas.

6.1.2 Gestión de Continuidad en las Contribuciones de Salida

El Gestor de Continuidad recibe las secuencias de actos comunicativos generadas por el Generador de Diálogo y coordina su adaptación multimodal y su generación de lenguaje natural. Esto permite que el resultado pueda ser integrado con la mayor continuidad y fluidez posible a la contribución en curso del sistema en los casos de reformulación. Los resultados de la gestión de continuidad son enviados a los Componentes de Interfaz de Salida, obteniendo como respuesta las notificaciones de síntesis de los fragmentos que la componen (a medida que el turno se desarrolla). Del mismo modo, este componente es el encargado de notificar el estado de actividad de turno del sistema al Gestor de Toma de Turno y de sintetizar los marcadores que son expresados como alteración en la continuidad temporal de las contribuciones [Figura 17].

El Gestor de Continuidad recurre a los generadores de lenguaje natural y al Adaptador Multimodal para generar una conjunto multimodal de expresiones naturales que se corresponden con la secuencia de actos comunicativos que el Generador de Diálogo obtuvo como resultado de una solicitud de formalización. Estas expresiones, sintetizadas de forma coordinada, hacen comprensible a los interlocutores la contribución que el sistema generó en forma de actos comunicativos. Así, por ejemplo, el sistema podría expresar como "*Tiene reunión con su director de tesis a las 12 y reunión de proyecto a las 5:30*" la secuencia de actos comunicativos generados por el Gestor de Diálogo *[< Informar de reunión con su director de tesis a las 12 ><Informar de reunión de proyecto a las 5:30>]* (que pudo ser generado como respuesta a lo que el sistema interpretó como la contribución completa de usuario "*lístame qué avisos tengo*").

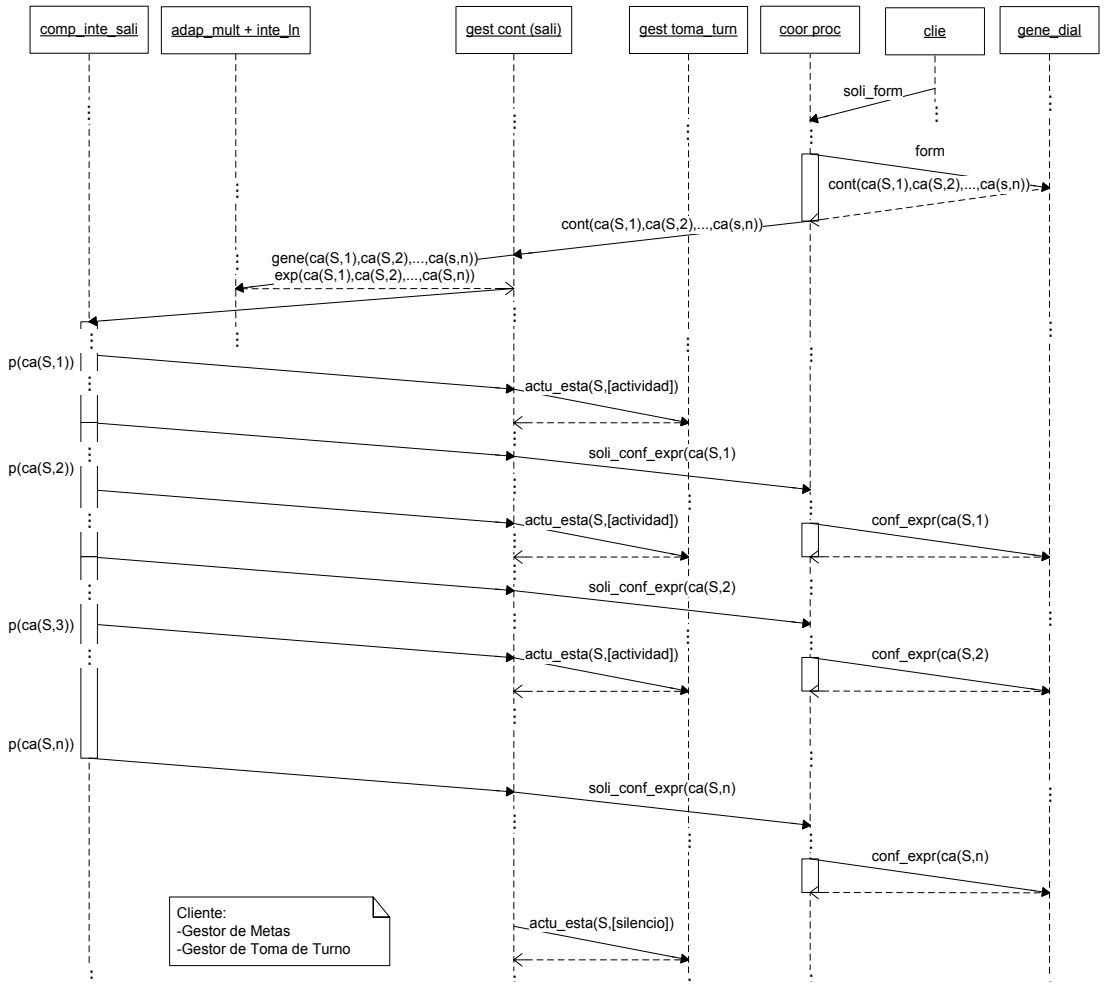


Figura 17: Diagrama de secuencia de gestión de continuidad de las contribuciones de salida

6.1.2.1 Generación Incremental

El conjunto multimodal de expresiones de lenguaje natural es remitido a los Componentes de Interfaz de Salida para que gestionen su síntesis. A medida éstos que van siendo sintetizados por el canal, los Componentes de Interfaz de Salida envían notificaciones al Gestor de Continuidad, quien comprueba si han sido expresados todos los fragmentos correspondientes al acto comunicativo en curso. Cuando esto sucede, desencadena, a través del Coordinador de Procesos, la confirmación de expresión del acto comunicativo en el Gestor de Diálogo. En el caso del ejemplo, suponiendo por sencillez granularidad de palabra, el Coordinador de Procesos confirmaría el acto comunicativo <*Informar de reunión con su director de tesis a las 12* > tras ser notificada la expresión del fragmento "*12*", y el acto <*Informar de reunión de proyecto a las 5:30*> tras la notificación de "*5:30*".

Al ser tratada la generación de forma incremental, el Generador de Diálogo podría reformular su contribución en curso como consecuencia de cambios en el estado de las metas o en el estado de la toma de turno. Si en el ejemplo el sistema, mientras está expresando su contribución "*Tiene reunión con su director de tesis a las 12 y reunión de proyecto a las 5:30*", recibe los nuevos fragmentos de contribución de su interlocutor "*para*" y "*mañana*", podría reinterpretar como "*lístame qué avisos tengo mañana*" lo que inicialmente interpretó como "*lístame qué avisos tengo*". De esta forma, el Generador de Diálogo podría reformular su contribución, obteniendo como resultado "*Tiene clase a las 11*" (en lugar de lo que inicialmente estaba expresando).

Es de esperar que, para cuando una contribución se reformula, parte de su formalización previa haya sido ya expresada por el canal. En estos casos debe ser gestionada la forma en la que la nueva formalización se combina con la porción de contribución previamente expresada para que la transición se produzca con la mayor continuidad y fluidez posible. El Gestor de Continuidad es el componente encargado de llevar a cabo este proceso. Por ello, monitoriza cuál de los actos comunicativos de la contribución está siendo sintetizado por la interfaz en cada momento, y qué porción de sus fragmentos de expresión ha sido ya expresada. Al ser recibida la reformulación, el Gestor de Continuidad identifica si existe un *punto de transición* entre la formalización previa y la nueva y, de ser así, cuál es. En el caso del ejemplo, si para cuando el Gestor de Continuidad recibe la reformulación "*Tiene clase a las 11*" el sistema ya había sintetizado el fragmento de contribución "Tiene" (de su formalización previa), sólo le quedaría por expresar "*...clase a las 11*", con lo que la contribución resultante quedaría "*Tiene | clase a las 11*" (dónde "|" denota el punto de transición).

Se considera que existe un punto de transición entre la formalización previa y la reformulación si la nueva comienza por un *acto comunicativo compatible* con alguno de los actos comunicativos de la formulación previa. Dos actos comunicativos son compatibles si pueden ser expresados en lenguaje natural según patrones que encajen en algún punto. De esta forma, si el acto comunicativo <*Informar de reunión con su director de tesis a las 12* > tiene asociado el patrón de lenguaje natural "*Tiene reunión con su director de tesis a las 12*", los actos comunicativos <*Informar de clase a las 11* > y <*Informar de reunión a las 11*> (teniendo asociados, respectivamente, los patrones "*Tiene clase a las 11*" y "*Tiene reunión a las 11*") serán compatibles con él, pero no lo será <*Informar de que no hay avisos*> (teniendo asociado el patrón "*No tiene avisos*") [Tabla 8].

Tabla 8: Ejemplos de actos comunicativos y sus relaciones de compatibilidad

Fragmento compatible	**AC_4**	**AC_3**	**AC_2**	**AC_1**
AC_1	No compatible	*"Tiene reunión"*	*"Tiene"*	*"Tiene reunión con su director de tesis a las 12"*
AC_2	No compatible	*"Tiene"*	*"Tiene clase a las 11"*	
AC_3	*No compatible*	*"Tiene reunión a las 11"*		
AC_4	"No tiene avisos"			

Acto Comunicativo	**Expresión en Lenguaje Natural**
AC_1: *Informar de reunión con su director de tesis a las 12*	*"Tiene reunión con su director de tesis a las 12"*
AC_2: *Informar de clase a las 11*	*"Tiene clase a las 11"*
AC_3: *Informar de reunión a las 11*	*"Tiene reunión a las 11"*
AC_4: *Informar de que no hay avisos*	"No tiene avisos"

Dado que un mismo acto comunicativo podría admitir representaciones alternativas en lenguaje natural (a través de una o varias modalidades), al ser solicitada la expresión multimodal y en lenguaje natural del acto comunicativo sobre el que se realizará la transición, se adjuntará a dicha solicitud la porción de expresión del acto comunicativo ya sintetizada, para permitir al Adaptador Multimodal y a los generadores de lenguaje natural elegir, de todos los posibles patrones, el que permita una mayor continuidad de la contribución resultante.

Una vez diseñada la expresión multimodal y en lenguaje natural de la reformulación (habiendo considerado las restricciones pertinentes para la generación del acto comunicativo compatible), es identificado el punto de transición entre las contribuciones. La porción de contribución restante (la comprendida entre el punto de transición y el final de la reformulación de la contribución), es remitida a los Componentes de Interfaz de Salida quienes, habiendo procedido a la cancelación de la expresión en curso de la contribución previa (tal y como define

la estrategia de coordinación de procesos de la propuesta), continuarán con la síntesis de la contribución (de acuerdo a la versión reformulada de la contribución).

En función de que exista o no un punto de transición entre la formalización previa y la reformulación, podrá distinguirse entre reformulación suave, rectificación y auto interrupción.

Reformulación Suave

Ambas formalizaciones encajan y el punto de transición es posterior a la porción expresada hasta el momento: En este caso, puede producirse una transición suave entre formulaciones [Ejemplo 18].

Formalización previa: "*hay una alarma para dentro de dos minutos*"
Reformulación: "*hay una alarma para dentro de un minuto*"
Contribución expresada: "*hay una alarma para dentro de | un minuto*"

Ejemplo 18: Reformulación suave

Generalmente este tipo de reformulación se produce de forma imperceptible para los interlocutores, aunque en ocasiones podría requerir una pausa en el desarrollo de la contribución (si la necesidad de reformular se detectó a tiempo, pero es imposible tener la reformulación preparada para el instante en el que estaba previsto que fuera expresada). Ante estas situaciones, pueden incluirse en la contribución del sistema carraspeos y otros tipos de silencios oralizados como estrategias de retención de la palabra ("hay *una alarma para dentro de… un minuto*").

Rectificación

Ambas formalizaciones encajan, pero el punto de transición ya fue expresado. En estos casos deben desarrollarse estrategias de rectificación. Cuando la reformulación de la contribución del sistema afecta a la porción del discurso ya expresada, deberán incorporarse técnicas de rectificación en el desarrollo de la contribución. En función del alcance de la rectificación, pueden aplicarse distintos tipos de reparación [32, pp.258-266]:

- *Reemplazamiento instantáneo*: Sólo reemplaza el elemento a reparar y viene asociado a una detección temprana ("*hay una alarma para dentro de do…| un minuto*").
- *Reemplazamiento completo*: Cuando no se suspendió el desarrollo inmediatamente después al elemento a reparar, es necesario reemplazar el elemento a reparar y todos los elementos que lo siguieron *(*"*hay una alarma para dentro de dos min| un minuto*").

- *Reemplazamiento anticipatorio*: Para facilitar la identificación del elemento a reparar (cuando el compromiso está debilitado) es posible incluir en ella algunos elementos anteriores ("*hay una alarma para dentro de dos min| de un minuto*").
- *Comenzar de nuevo*: Consiste en abandonan la presentación completamente para empezar de nuevo desde el principio. Suele estar relacionada con rectificaciones profundas de la contribución ("*Para dentro de dos minutos hay program…| Para dentro de un minuto hay programada una alarma*").

A menudo la reparación se acompaña de silencios que permiten identificar el punto de reparación.

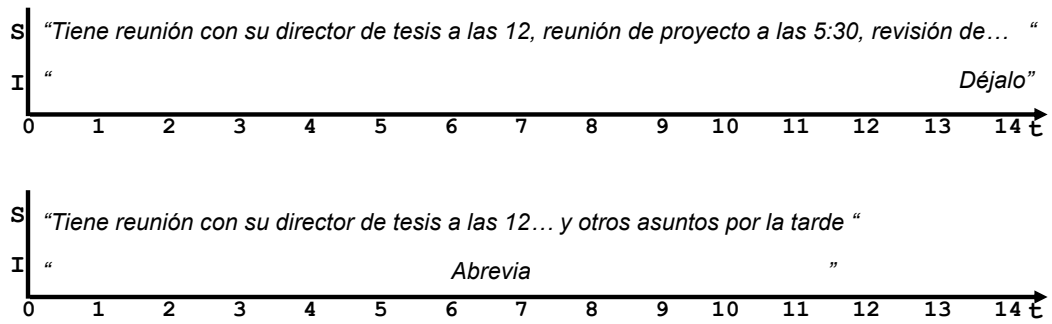
Auto interrupción

Si las formalizaciones no encajan, el sistema se encuentra ante una auto interrupción. La variación de la prioridad de unas metas sobre otras puede hacer preferible desestimar la contribución que el sistema ha formalizado y está expresando para sustituirla por una nueva. En la medida en la que la urgencia de las nuevas metas a desarrollar sea elevada, el Gestor de Continuidad podría recibir reformulaciones incompatibles con la porción de contribución expresada hasta el momento, de forma que no pueda ser gestionada la continuidad de la contribución. En estos casos se produce una ruptura brusca en el desarrollo de la contribución que en muchos casos puede ser suavizada con intervalos de anuncio y con posteriores reparaciones (cuando la urgencia de las metas ha sido controlada) [Ejemplo 19].

Formalización previa: "*Tiene reunión con su director de tesis a las 12 y reunión de proyecto a las 5:30*"
Reformulación: "*¡Por cierto!, tiene cita con el médico para dentro de media hora*"
Contribución expresada: "*Tiene reunión con su director de tesis a las 12 y reunión de proyecto a las … ¡Por cierto!, tiene cita en el médico para dentro de media hora*"

Ejemplo 19: Auto interrupción con anuncio

Finalmente, los cambios producidos en las metas de la interacción o en el estado de la toma de turno podrían llevar al sistema a interrumpir o finalizar prematuramente el desarrollo de su propio turno. En tales casos, la nueva formalización consistirá en una contribución nula o una breve contribución de desenlace [Ejemplo 20].



Ejemplo 20: Interrupción y finalización prematura de la contribución del sistema

6.1.2.2 Notificación de Actividad e Inactividad

Entre las funciones del Gestor de Toma de Turno se encuentra la de estimar el estado en que se encuentran los turnos de los distintos participantes de la interacción. Sólo de esta forma será posible conjeturar el estado en el que se encuentra la posesión de la palabra en la interacción y quiénes son los posibles candidatos a tomarla. Esto incluye también la representación del turno del propio sistema, para el que la actualización del estado de actividad corresponde al Gestor de Continuidad (al igual que sucede con los turnos del resto de participantes).

Aunque el sistema conoce con total certeza lo que ocurre con su propio turno (sabe cuando comienza su actividad y cuando cesa), por lo que en realidad se rige el reparto de turnos de la interacción (considerándolo una actividad combinada), no son las actividades individuales que desarrolla cada participante, sino las estimaciones que todos ellos puede realizar sobre lo que piensan los demás. Sólo de este modo podrán alcanzar una representación común del estado en el que se encuentran los turnos, la posesión de la palabra y los candidatos a tomarla. Por tanto, el interés no es representar el estado real en el que se encuentra el turno del sistema (la actividad individual del sistema), sino lo qué pueden conocer y estimar el resto de participantes sobre él. De esta forma, las notificaciones de actividad y cese de actividad del turno del sistema serán generadas siguiendo las mismas reglas que las de los turnos del resto de participantes.

El sistema, al igual que cualquier otro interlocutor, tendrá asociado su propio autómata para determinar las situaciones en las que el Gestor de Toma de Turno debe ser informado de la actividad del turno del sistema. El estado de dicho autómata será actualizado, como ocurre con el resto de participantes, cada vez que el Gestor de Continuidad detecte que el sistema ha expresado un nuevo fragmento de contribución (cuando recibe de los Componentes de Interfaz de Salida confirmaciones de síntesis). Del mismo modo, será informado ante silencios prolongados en el turno del sistema o silencios oralizados.

6.1.2.3 Marcadores de Gestión de la Toma de Turno

Al igual que el resto de participantes, el sistema también puede desarrollar marcadores de gestión de la toma de turno basados en la alteración de la continuidad de la contribución (en cuyo caso es competencia del Gestor de Continuidad sintetizarlos). Estos marcadores están relacionados con la solicitud de palabra, su mantenimiento o su cesión Tal y como fue explicado en el Apartado 6.1.1.3 son, entre otros, *retardo momentáneo del turno*, *reinicio del turno*, *parada y continuación* o *comienzo y parada*.

6.2 *COORDINADOR DE PROCESOS*

El principio de secuencialidad supuesto tradicionalmente en el desarrollo de las distintas fases del *ciclo de interacción* (adquisición, interpretación, operación, generación y síntesis) queda roto bajo la condición de toma de turno no marcada. Tal flexibilidad en la toma de turnos requiere del sistema la capacidad para gestionar simultáneamente contribuciones de distintos participantes a la vez, incluidas las del propio sistema, lo que supone un acceso concurrente a un conjunto de recursos compartidos relacionados con el conocimiento sobre la interacción y sobre las circunstancias que la rodean. Estas situaciones de contribución simultánea pueden ocurrir como consecuencia de la comunicación colateral simultánea, por solapamiento en las transiciones entre hablantes, durante la resolución de interrupciones, por error, etc. La coordinación de procesos es la estrategia según la cuál se controla el acceso de los distintos procesos a los recursos compartidos. Debe producirse garantizando:

- Que el acceso a estos recursos sea producido en exclusiva, puesto que se trata de recursos no compatibles que requieren ser consultados y actualizados por todos los procesos y que, de ser usados de forma concurrente por varios de ellos, podrían alcanzar estados inconsistentes.
- Que se evite la monopolización de los recursos por alguno de los procesos, al ser indispensable que todos ellos sean atendidos en tiempo real (según van ocurriendo) para desarrollar una toma de turnos no marcada. De otro modo sería imposible gestionar la realimentación; la actualización de las contribuciones del sistema como efecto de cambios de las circunstancias sociolingüísticas; el desarrollo de participaciones solapadas e interrupciones; etc.
- Que exista una coherencia entre las representaciones internas que mantienen los distintos participantes sobre el estado de la interacción y sobre el resto de

conocimiento sociolingüístico que la rodea. Ello requiriere que todos los procesos actualicen y apliquen todo este conocimiento en el orden correcto.

6.2.1 Control de Acceso a los Recursos Compartidos

La interpretación de la contribución completa de un interlocutor será tramitada como una secuencia de interpretaciones incrementales desarrolladas puntualmente a intervalos dados por una determinada granularidad temporal. Cada vez que el Gestor de Continuidad identifica fragmentos de contribución que pueden ser interpretados a nivel de diálogo (actos comunicativos), requerirá su interpretación de diálogo. Por su parte, la generación también será cursada de forma incremental, lo que implica diferenciar entre la formalización inicial de una contribución, solicitada por el Gestor de Metas y el Gestor de Toma de Turno, y las confirmaciones parciales de su expresión, solicitadas por el Gestor de Continuidad. De esta forma, se identifican tres tipos distintos de proceso: *interpretación de actos comunicativos*; *formalización de nuevas contribuciones*; y *confirmación de expresión de actos comunicativos*. Todos ellos son procesos que requieren el acceso a determinados recursos compartidos, como son el Estado de Interacción y el conocimiento sociolingüístico asociado.

El Coordinador de Procesos será el componente encargado de planificar la forma en la que estos procesos acceden a los recursos compartidos. Para garantizar el acceso exclusivo a los mismos, se basa en un planificador cuya cola de recepción de solicitudes está organizada por niveles de prioridad [Figura 18]. El planificador extrae las solicitudes de forma secuencial y ordenada y las ejecuta de una en una (evitando el riesgo de acceso concurrente a los recursos).

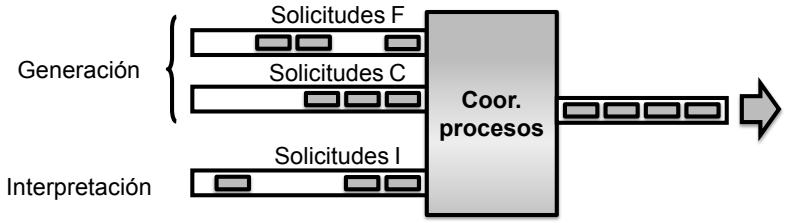


Figura 18: Coordinador de Procesos

6.2.2 Gestión de Esperas

Se requiere que todas las solicitudes recibidas por el Coordinador de Procesos sean atendidas en tiempo real. Dado que los procesos interpretación y generación (ambos de naturaleza continua) han sido descompuestos en subprocesos discretos (*interpretación de actos comunicativos*, *formalización de contribución* y *confirmación de expresión de acto*

comunicativo), y considerando que todos ellos tienen un tiempo de ejecución despreciable frente a los tiempos de producción del *lenguaje natural* [Apartado 5.1.1], podrá concluirse, de aplicar las Ecuaciones de Little para la Teoría de Colas [89], que todas serán atendidas dentro de un retardo tolerable para la *interacción natural*.

Para la demostración, se supondrá una distribución exponencial de la tasa de llegadas y del tiempo de servicio. Al ser atendidos los procesos de uno en uno, el Coordinador de Procesos podrá modelarse como una cola M/M/1, según la notación Kendall. El número previsto de solicitudes a la espera (L_q) viene dado para este caso por la Ecuación 1, que supuesto un tiempo de servicio ($1/\mu$) despreciable frente al tiempo entre llegadas ($1/\lambda$) permite determinar que las esperas que se prevén para la atención de solicitudes en el Coordinador de Procesos tenderán a cero.

$$\lim_{\frac{\mu}{\lambda}\to\infty} L_q = \lim_{\frac{\mu}{\lambda}\to\infty} \frac{\lambda^2}{\mu\cdot(\mu-\lambda)} = \lim_{\frac{\mu}{\lambda}\to\infty} \frac{1}{\frac{\mu}{\lambda}\cdot\left(\frac{\mu}{\lambda}-1\right)} = 0$$

Ecuación 1: Cálculo del número previsto de solicitudes a la espera en el Coordinador de Procesos

Si bien es cierto que la suposición de partida (tiempo de ejecución despreciable frente a la llegada de nuevas solicitudes) es muy dependiente de las características hardware de las máquinas sobre las que se despliegue la plataforma y de la implementación de los componentes, lo que se busca en este trabajo es la definición de un modelo que permita desarrollar una toma de turno avanzada, sin tener en cuenta cuestiones relacionadas con la eficiencia de la implementación o de las máquinas sobre las que se despliega. En cualquier caso, los sistemas informáticos actuales permiten obtener tiempos de procesamiento suficientemente, bajos con respecto a los tiempos de producción del lenguaje natural, como para dar por válidas dichas suposiciones. Aun en el caso de que esto no pudiera darse por cierto, la solución pasaría por reajustar el valor de la granularidad de procesamiento temporal aplicada. Una granularidad temporal demasiado alto llevaría a una pérdida de la inmediatez con la que los participantes reciben las evidencias de cierre a sus acciones, pero la respuesta seguiría siendo la mejor posible dada las restricciones físicas. De esta forma, puede concluirse que no se requieren, en ningún caso, estrategias adicionales para el desalojo de los procesos del coordinador.

6.2.3 Orden de Ejecución de los Procesos

El orden que establece la cola de solicitudes para los casos de colisión se fundamenta en la necesidad de que las estimaciones sobre el estado de conocimiento que alcanzan tanto el

sistema como sus interlocutores sean similares. De acuerdo a este criterio, las solicitudes serán procesadas en el siguiente orden:

1. Confirmaciones de expresión de actos comunicativos del sistema (C): Provienen del Gestor de Continuidad y serán ejecutadas por el Gestor de Diálogo, tal y como muestra la Figura 19.

2. Interpretación de nuevos actos comunicativos de los interlocutores (I): Estas solicitudes provienen del Gestor de Continuidad y son ejecutadas por el Intérprete de Diálogo, como aparece reflejado en la Figura 20.

3. Formalización de nuevas contribuciones del sistema (F): Desencadenadas por el Gestor de Toma de Turno o el Gestor de Metas ante los cambios en el estado de las metas; los cambios en el estado de los turnos de los participantes; los cambios en la posesión de la palabra; o los cambios en el conjunto de candidatos a tomarla [Figura 21].

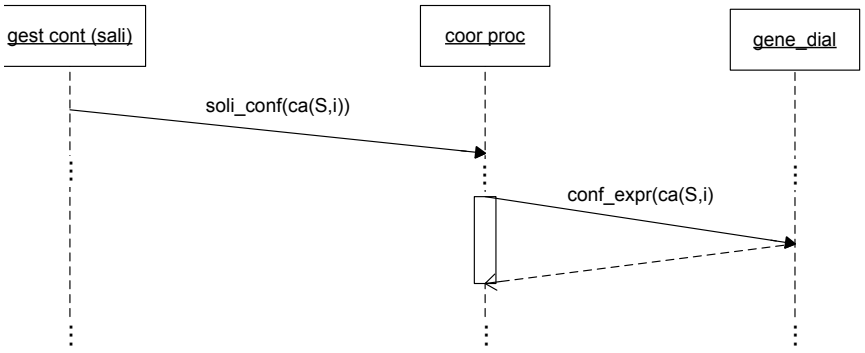


Figura 19: Diagrama de secuencia de la coordinación de procesos en la confirmación de expresión de un acto comunicativo.

En los casos de recepción de nuevas solicitudes de interpretación de actos comunicativos y de formalización de nuevas contribuciones, el Coordinador de Procesos deberá considerar, adicionalmente, si el sistema está desarrollando simultáneamente un proceso de generación (el sistema está tomando turno para desarrollar una contribución). En estos casos también deberá gestionar su cancelación. Ante dichas situaciones, el Coordinador de Procesos solicitará, en primer lugar, la cancelación de la contribución en curso del sistema a los Componentes de Interfaz de Salida (a través del Gestor de Continuidad). Una vez confirmada la cancelación por los Componentes de Interfaz de Salida, se procederá a ejecutar las confirmaciones de expresión que pudieran aun estar pendientes (algunas de las cuales podrían

haber sido producidas durante la gestión de la cancelación en los Componentes de Interfaz de Salida). Finalmente, se ejecutará la interpretación o formalización que provocó la cancelación.

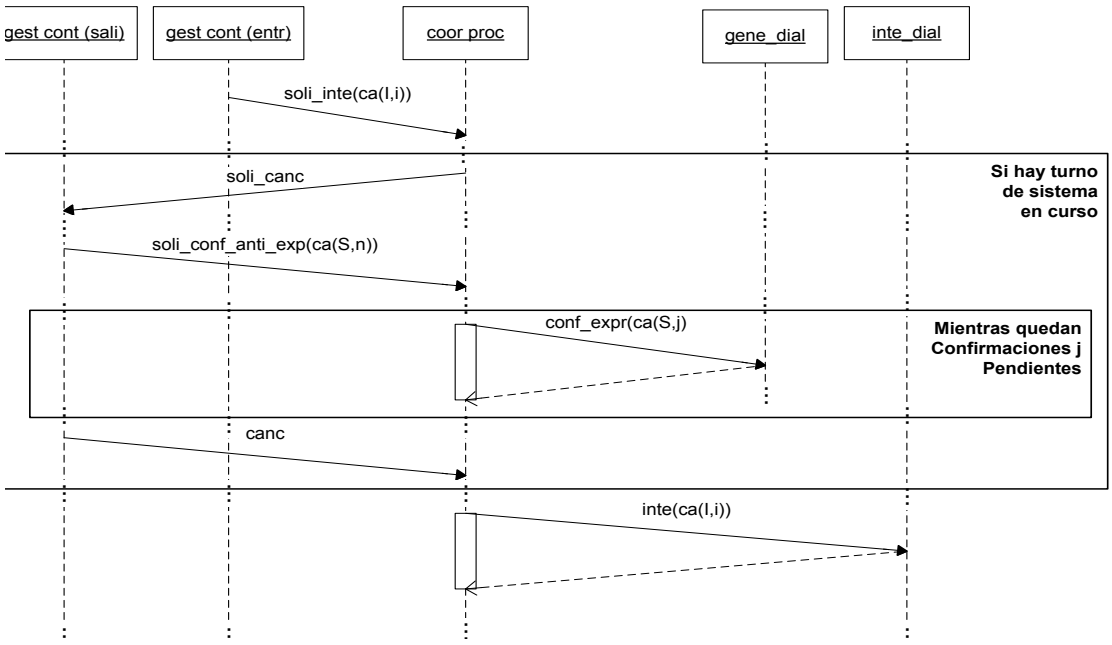


Figura 20: Diagrama de secuencia de la coordinación de procesos en la interpretación de actos comunicativos

En el caso de una cancelación por solicitud de interpretación, si las razones que llevaron al sistema a formalizar la contribución que se canceló siguieran estando vigentes, el Gestor de Toma de Turnos o el Gestor de Metas desencadenarían una nueva formalización de contribución que, al ser atendida, resultaría en la continuación de la contribución cancelada. Si la cancelación fue provocada por una solicitud de formalización, en la contribución resultante sería también considerada la continuación de la contribución previa. En ambos casos, en función de que las razones que llevaron a la generación de la contribución cancelada hayan sido alteradas en mayor o menor medida (como consecuencia de la ejecución de dichos procesos), el resultado será una reformulación, una rectificación, o la auto interrupción de la contribución. En cualquiera de ellos, este mecanismo permite actualizar la contribución en curso del sistema con los nuevos eventos ocurridos simultáneamente a su generación (bien sea la contribución simultánea de sus interlocutores; el cambio en el estado de los turnos, la posesión de la palabra o los candidatos; o el cambio en el estado de las metas de la interacción).

Sobre ejemplo de contribución formalizada por el sistema "*Tiene una reunión de tesis a las 12, reunión de proyecto a las 5:30 y clase mañana a las 11:00*" (producida como consecuencia de una solicitud de su interlocutor de un listado de los avisos pendientes), la

gestión de coordinación propuesta permitiría desarrollar simultáneamente los procesos involucrados en la interacción del Ejemplo 21.

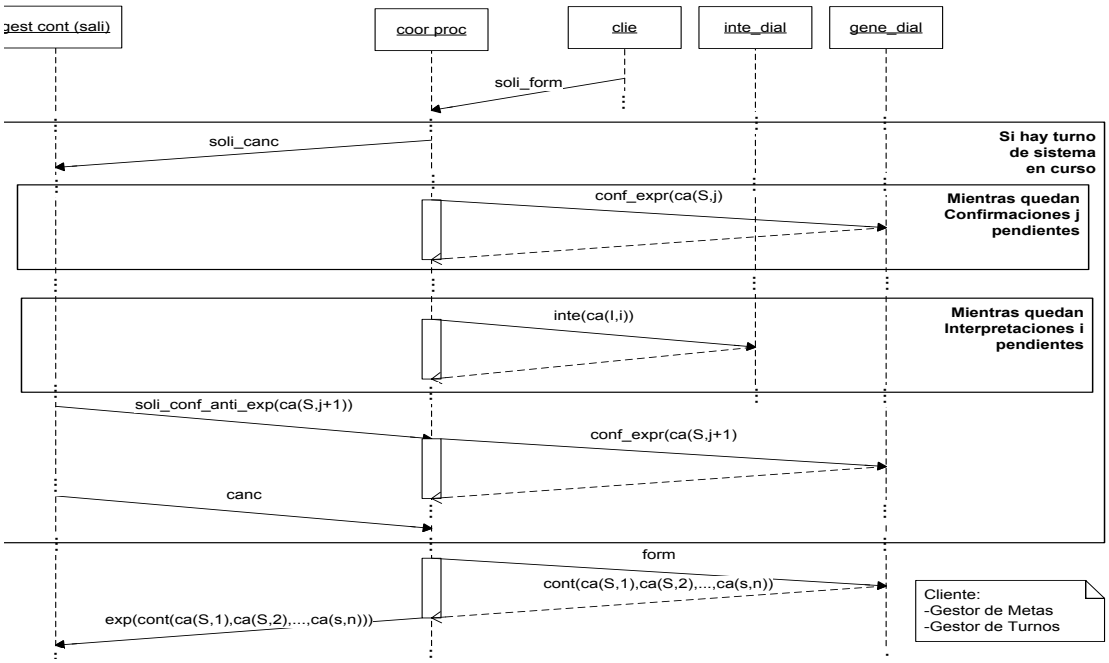
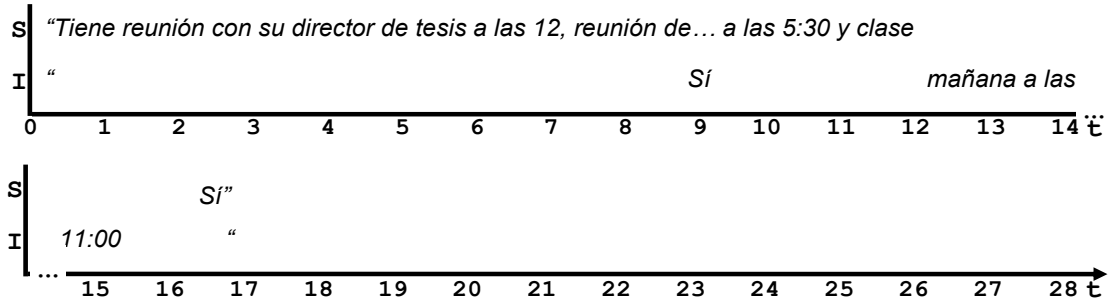


Figura 21: Diagrama de secuencia de la coordinación de procesos en la formalización de una nueva contribución

Tanto las contribuciones de usuario, como las de sistema, son procesadas a nivel de fragmento de contribución. Supuesta una granularidad de palabra, esto querrá decir que el sistema recibe notificaciones cada vez que una palabra ha sido sintetizada y que podrá confirmar los progresos desencadenados por los actos comunicativos tan pronto como se haya notificado la expresión de todas las palabras que expresan su acción comunicativa. Del mismo modo, se notifica la adquisición de nuevos fragmentos de contribución con cada palabra y se ejecuta un proceso de interpretación incremental cada vez que se han recibido suficientes palabras como para constituir por sí mismas una acción comunicativa. La confirmación de la expresión de los actos comunicativos del sistema y la interpretación de diálogo de los nuevos actos comunicativos adquiridos de su interlocutor serán las solicitudes cuya coordinación se detalla en el Ejemplo 21.

La enumeración inicialmente formalizada por el sistema podría ser estructurada en tres acciones comunicativas distintas: <*informar de una reunión con el director de tesis a las 12*> (AC_1); <*informar de una reunión de proyecto a las 5:30*> (AC_2); y <*informar de clase mañana a las 11*> (AC3). Esta contribución comienza a ser desarrollada sobre la formalización inicial y, a medida que van siendo expresados los fragmentos de contribución correspondientes a actos

comunicativos, el sistema puede ir tramitando la confirmación de los progresos asociados a cada uno de ellos. Así, en el momento en que el interlocutor enuncia "*Sí*" (contribución U_1, expresada en 16<t<17 y que, por si misma, constituye un acto comunicativo), el sistema habría confirmado el progreso correspondiente a AC_1 (puesto que habría expresado todas las palabras en que se sintetiza), pero no así el correspondiente a AC_2 (del que sólo ha sido expresado "*reunión de*"). Tampoco podría darse por confirmado AC_3 (del que aún no ha comenzado a expresarse nada). La contribución U_1 ofrece realimentación y permite ajustar el nivel de información aportado por el sistema sobre la cita, provocando la reformulación de la porción de contribución pendiente. En consecuencia se formaliza la nueva secuencia de actos comunicativos $AC_{2'}$, con menor nivel de información ("*informar de una reunión a las 5:30*"), y AC_3, que no sufre variaciones. Tras ser encadenada convenientemente por el Gestor de Continuidad con la porción de intervención cursada hasta el momento, daría lugar a la reformulación S_2, expresada por el sistema en 9<t<12. Cabe destacar que, según el carácter de U_1 (desvíe o no la atención, pueda resultar agresiva, etc.), el compromiso podría verse afectado y, en consecuencia, sería necesario aplicar estrategias de refuerzo (por ejemplo redundancia: "*Esta reunión es a las doce*").



Ejemplo 21: Efecto de la coordinación de procesos sobre la interacción natural

En 12<t<15 el interlocutor produce la contribución "*mañana a las 11*" (U_2, con entidad como acción comunicativa) y, con ello $AC_{2'}$ habría sido confirmada por completo y sólo quedaría pendiente de confirmación AC_3 (de la que sólo fue expresado "*y clase*"). La contribución U_2 constituye una interrupción colaborativa y desencadena la cancelación del desarrollo del resto de contribución del sistema. La contribución producida por el sistema en 16<t<17 ofrece al usuario las evidencias de que la continuación propuesta a la contribución previa del sistema fue correcta.

Si la coordinación de procesos consistiese en el bloqueo de los recursos compartidos por un único proceso durante su desarrollo completo, la contribución U_1 de usuario no habría sido procesada hasta haber concluido la formalización preliminar de la contribución del sistema y, por tanto, no habría sido posible desarrollar la reformulación S_2 ni su posterior interrupción.

Finalmente, con el objetivo de favorecer una transición fluida y continuada de la contribución en los casos de reformulación, los Componentes de Interfaz de Salida podrán confirmar prematuramente algunos de los fragmentos de la contribución, con el objetivo de expresarlos durante la gestión de la reformulación y evitar discontinuidades. Ante las solicitudes de cancelación, los Componentes de Interfaz de Salida confirmarán tantos fragmentos de contribución como se requiera para asegurar que, en caso de darse la reformulación, haya tiempo suficiente como para ejecutar todos los procesos implicados sin que cese el flujo de expresión por el canal. En el ejemplo de la Figura 22 se observa cómo es gestionada la confirmación anticipada del fragmento de expresión "*exp(ca(S,i-1))*", previo al fragmento inicial de la porción reformulada "*exp(ca(S,i))*". Si la granularidad temporal elegida para fragmentar las contribuciones es adecuada a las características hardware y software del sistema, sólo será preciso realizar la confirmación prematura de uno de los fragmentos de contribución.

El Ejemplo 22 muestra cómo la confirmación prematura de fragmentos de contribución permite asegurar la continuidad de la contribución del sistema. A medida que se desarrolla la contribución, debido a la interpretación simultánea de las contribuciones de realimentación de su interlocutor (producidas en los instantes t=6, t=10 y t=18), son desencadenadas cancelaciones de la contribución en curso (en los instantes t=7, t=11 y t=20, respectivamente) y, posteriormente, reformuladas (en t=8, t=12 y t=21). La confirmación prematura de los fragmentos de expresión que se sintetizarán durante la gestión de las reformulaciones (fragmentos "*tarde*", "*rodilla*" y "*caíste*", suponiendo granularidad de palabra) permiten evitar discontinuidades en el turno finalmente desarrollado por el sistema.

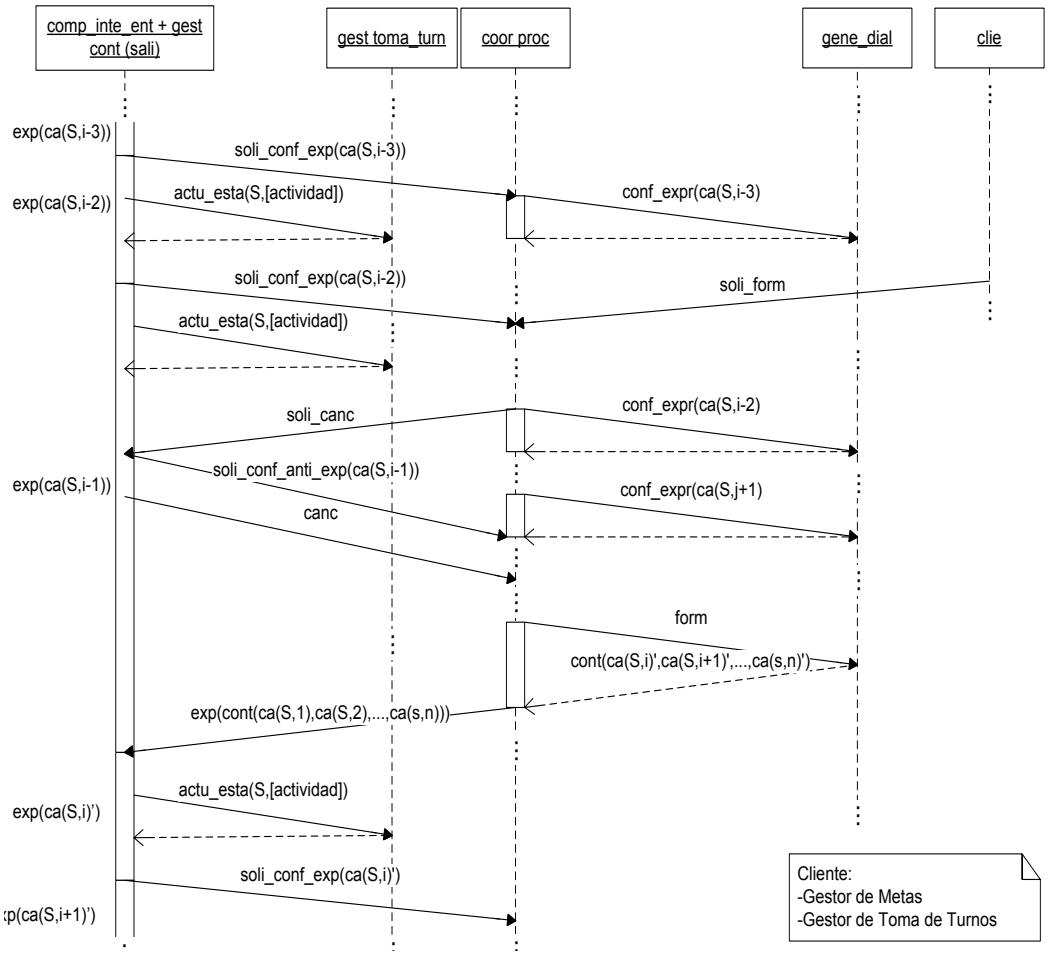
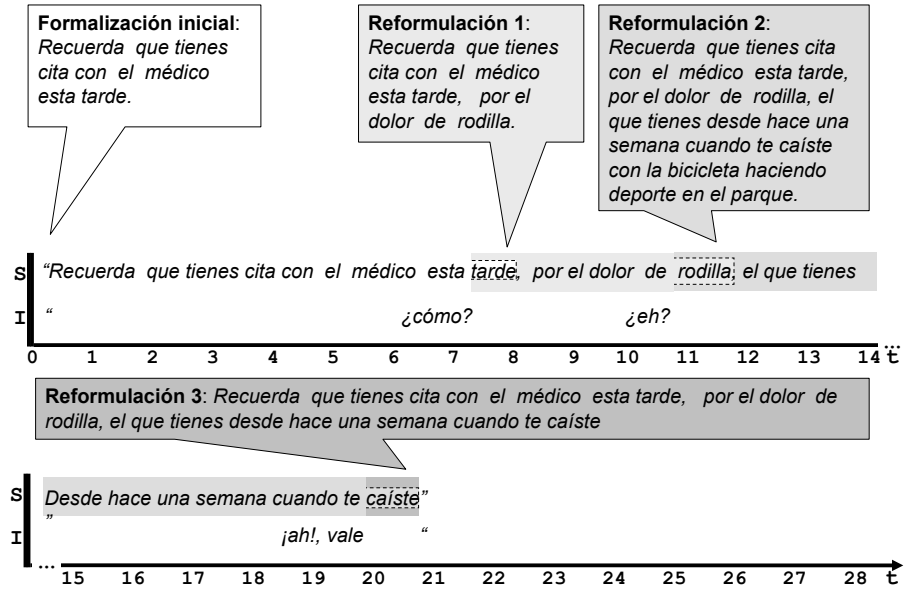


Figura 22: Gestión de la reformulación con confirmación anticipada de fragmentos de contribución



Ejemplo 22: Reformulación con confirmación prematura de fragmentos de contribución para evitar discontinuidades en la contribución resultante

6.3 GESTOR DE TOMA DE TURNO

El Gestor de Toma de Turno es el componente que representa las conjeturas que realiza el sistema acerca del estado en que se encuentra el turno de cada uno de los participantes, la posesión de la palabra y los candidatos a tomarla. Estas conjeturas van modificándose a partir de los marcadores detectados por el Gestor de Continuidad y Gestor de Diálogo a medida que progresan las interpretaciones de las contribuciones de los interlocutores y la generación de la del propio sistema. Del mismo modo, en función de los cambios que se producen en el estado de posesión de la palabra y en el de las metas (modeladas por el Gestor de Metas), pueden desencadenarse decisiones de toma de turno que serán desarrolladas por el componente Gestor de Toma de Turno (y podrían suponer nuevas solicitudes de formalización de contribución al Generador de Diálogo a partir del Coordinador de Procesos [Figura 23]).

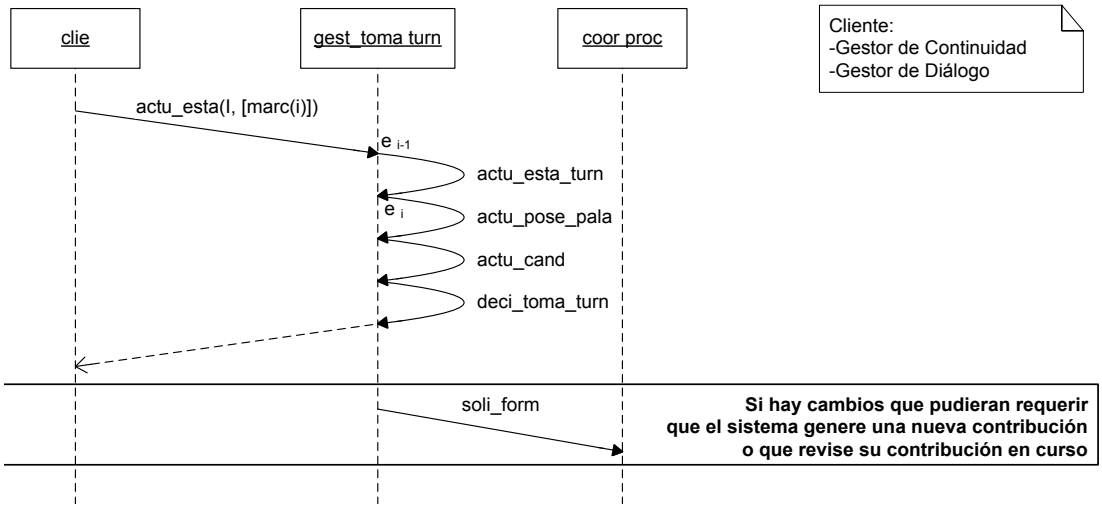


Figura 23: Diagrama de secuencia de la actualización del estado de la toma de turno

6.3.1 Estado de Turnos

A lo largo de la interacción es preciso identificar qué participantes se encuentran desarrollando turno en la interacción (participando en ella), detectar en qué momentos los participantes concluyen explícitamente sus turnos, y estimar cuándo podrían darse por concluidos (aun sin existir una expresión explicita de ello). Además, en el caso de los turnos que se desarrollen bajo la posesión de la palabra, será especialmente valiosa la proyección prematura de los puntos en los que sería tolerable un cambio de hablante, bien cuando el sistema opta a tomar la palabra para determinar los momentos en que puede intentar hacerse con ella, o bien cuando él es el hablante, para aceptar de buen grado el cambio de hablante en caso de ocurrir o anticipar medidas que le permitan evitarlo (gestos como levantar la mano o expresiones verbales

como "*un momento*"). Todas estas conjeturas se enmarcan dentro de la estimación del estado de los turnos.

El Gestor de Toma de Turno representa el estado del turno de cada participante. Esto es modelado a partir de una máquina de estados y de un indicador de su *pista de acción* [Apartado 2.2.3.2]. La máquina de estados permite estimar el estado de actividad del turno de cada participante, así como los posibles *TRPs* que ocurren en él (sus posibles finales). Por su parte, el indicador de pista representa las conjeturas que realiza el sistema sobre la pista (primaria o secundaria) que se encuentra desarrollando el participante en su turno en cada momento. El Gestor de Toma de Turno recibe de los componentes Gestor de Continuidad y Gestor de Diálogo las solicitudes de actualización de la máquina de estados del turno de un participante [Figura 24] durante los procesos de interpretación (para los turnos de los interlocutores del sistema) y generación (para su propio turno). Las actualizaciones que realiza el Gestor de Continuidad sobre el estado del turno se producen cuando son recibidos, desde los Componentes de Interfaz de Salida, nuevos fragmentos de la contribución de un interlocutor, o cuando se notifica la expresión de los fragmentos de la contribución del turno del propio sistema. Por su parte, el Gestor de Diálogo también podrá actualizar la máquina de estados durante los procesos de interpretación de diálogo y de generación de diálogo, en la medida en la que se puedan deducir nuevos marcadores relacionados con la toma de turno a partir de los de los progresos producidos en la interacción.

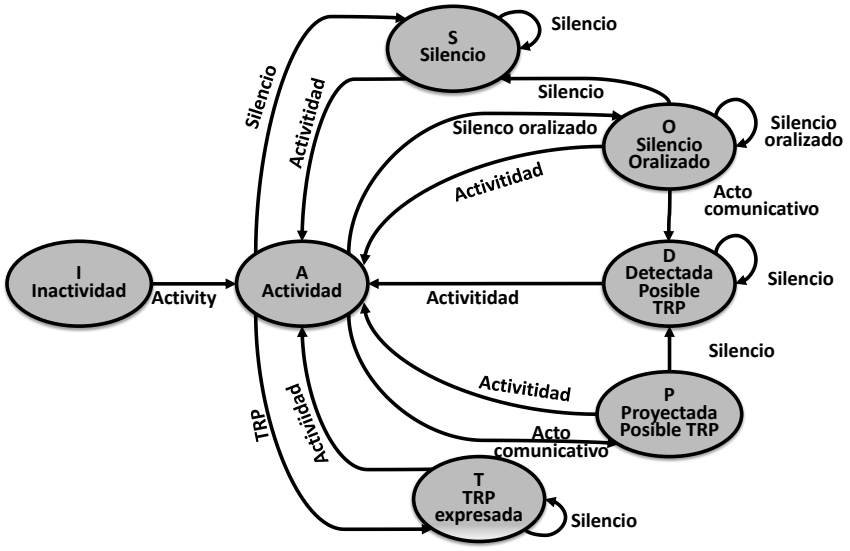


Figura 24: Máquina de estados del turno de un participante

Esta máquina de estados define los estados *inactividad* (*I*), *actividad* (*A*), *silencio* (*S*), *TRP expresada* (*T*), *detectada posible TRP* (*D*), *proyectada posible TRP* (*P*) o *silencio*

oralizado (*O*). El estado de *inactividad* (*I*) representa el estado inicial de turno o el estado alcanzado cuando el participante no está contribuyendo en la interacción (y, por tanto, no está tomando turno). El estado de *actividad* (*A*) representa un turno en desarrollo. *TRP expresada* (*T*) es el estado alcanzado cuando el participante que desarrolla un turno señaliza en él un final (con gestos o expresiones explicitas como, por ejemplo, "*no tengo nada más que decir*"). *Detectada posible TRP* (*D*) será el estado que tome el turno cuando el participante, tras haber expresado una contribución con significado interactivo completo (no existen fragmentos de expresión posteriores al último acto comunicativo interpretado), pase a estar en silencio (lo que podría suponer que el participante ha concluido su turno). *Proyectada posible TRP* (*P*) representará situaciones en las que podría anticiparse un posible final del turno (al estar próximo el final de acciones comunicativas completas en el turno del participante). Junto a estos, se incluyen en la máquina los estados *silencio* (*S*) y *silencio oralizado* (*O*) para representar discontinuidades en el desarrollo de un turno (producidas en forma de silencios o de silencios oralizados).

Como se recoge en la Figura 24, la transición entre el estado de *inactividad* (*I*) al estado de *actividad* (A) se produce cuando se recibe una notificación de actividad desde el Gestor de Continuidad. Ya en este estado de *actividad* (A), las posibles situaciones que podrían darse son: la ocurrencia de un silencio o de un silencio oralizado sin haber sido expresado un acto comunicativo completo (ambos detectados por el Gestor de Continuidad); la identificación de un acto comunicativo completo (lo que también es notificado por el Gestor de Continuidad, quien coordina a los componentes de Adaptación Multimodal y los procesadores de lenguaje natural); o la detección de una TRP expresada explícitamente (detectada por el Gestor de Diálogo). En el primero de los casos, la interrupción de la continuidad en la contribución no puede relacionarse con un posible final de la misma, por lo que se conjetura que el turno se encuentra realizando un *silencio* (*S*, lo que podría suponerse como una pausa) o un *silencio oralizado* (*O*, que estaría provocado por dificultades en la continuación de la contribución). En el segundo de los casos, el participante notifica explícitamente que su turno ha concluido, por lo que el turno pasaría automáticamente al estado final *TRP expresada* (*T*). Este estado sería corroborado en caso de recibirse silencios posteriores. Por su parte, la detección de actos comunicativos completos (lo que también podría ocurrir tras una serie de titubeos, repeticiones y, en definitiva, silencios oralizados) podría anticipar el final del turno, y debe ser considerada en el estado del turno para predecir en qué situaciones otro participante podría optar a tomar posesión de la palabra (de estar hablando el participante del cual se representa el turno). En estos casos, el estado será actualizado al de *proyectada posible TRP* (*P*). Si tras ello es notificada la ocurrencia de silencios, será reforzada la hipótesis de final del turno, con lo que el estado pasará al de *detectada posible TRP* (*D*).

En cualquiera de estas situaciones (aún cuando se detectó, estimó o proyecto una TRP), el Gestor de Continuidad podría notificar la continuación del desarrollo del turno del participante (con nuevos marcadores de actividad), con lo que la máquina de estados retomaría el estado de *actividad* (A). Del mismo modo, tras un periodo prolongado de tiempo sin ser recibidos marcadores en el turno por parte del participante, se ejecutará una transición al estado de *inactividad* (*I*), independientemente del estado en que se encontrase, dándose con ello finalizado el turno del participante.

Junto a la máquina de estados de cada participante se representa un indicador de la pista de acción en la que se desarrolla su turno. La pista de acción podrá tomar los valores *turno primario* o *secundario* y se actualiza cada vez que el Estado de Interacción notifique un cambio en el *hilo enfocado*, o cuando cambia el hilo que desarrolla el participante en su turno (Lo que es notificado por el Interprete o Generador de Diálogo, en función de que se trate del turno de un interlocutor o del turno del propio sistema). El indicador de pista de acción representa la relación existente entre la pista del hilo que desarrolla actualmente el participante en su turno y la del hilo enfocado. Las pistas asociadas a los hilos de la interacción guardan entre ellas una relación jerárquica. De esta forma, cada pista puede tener, a su vez, una pista padre (a la que es secundaria) y una serie de pistas hijas (de las que es primaria). Cuando la pista del hilo que desarrolla el participante es la misma que la del hilo enfocado (o alguna de sus ancestros), el turno será considerado primario en la interacción (y secundario en cualquier otro caso). De esta forma, a medida que evoluciona el foco de la interacción y la contribución del participante, el carácter primario o secundario de su turno podrá cambiar, pudiendo pasar un turno de primario o secundario (o viceversa) a medida que transcurre la interacción.

Ante cualquier cambio en el estado del turno de los participantes, el estado de posesión de la palabra será revisado, lo que permitirá identificar situaciones en las que deberán cursarse solicitudes al Generador de Diálogo (a través del Coordinador de Procesos) para dar al sistema la oportunidad de reformular o interrumpir su contribución en curso, o de evaluar si debe formular una nueva. Este es el mecanismo a través del cual el sistema podrá resolver algunos fenómenos relacionados con la toma de turno, como son el arrebatar la palabra al hablante cuando no la está utilizando de forma efectiva (desarrollo injustificado de pausas o pausas oralizadas), o el adelantarse en las detecciones de TRPs, posibles TRPs, o incluso en su proyección para producir un nuevo turno (con el objetivo de evitar que otros participantes puedan adelantarse a tomar la palabra cuando la criticidad de sus propias metas o el compromiso de las metas combinadas lo requiere).

6.3.2 Estado de Posesión de la Palabra

La toma de turno de la interacción humana se caracteriza por:

- No quedar restringido al hablante el derecho a tomar el turno, puesto que muchos casos de solapamiento están justificados y son necesarios para una correcta gestión de la interacción. Tal es el caso de las contribuciones desarrolladas en la pista secundaria (que pueden expresarse en cualquier momento siempre que no desvíen de forma significativa la atención sobre la presentación primaria), o cuando se va a producir un cambio de hablante (cuando en ocasiones el hablante siguiente adelanta el comienzo de la producción de su contribución para evitar que otros participantes puedan arrebatarle la palabra).
- No estar justificada la toma de turno en cualquier caso. A pesar existir situaciones que hacen posible que un participante desarrolle un turno sin estar en posesión de la palabra, la mayoría de las situaciones hace preferible ser el hablante legítimo para desarrollar un turno. Lo contrario sería mal recibido por los interlocutores y la salud de la interacción se resentiría. No obstante, queda en manos de los participantes la decisión final de violar o no las reglas de toma de turno (si a su juicio tal violación compensa los daños causados).
- No quedar garantizado que las circunstancias sociolingüísticas, el estado de interacción y el estado de los turnos que justificaron la toma de turno sean válidos durante el desarrollo de toda la contribución. A medida que las condiciones en las que se desarrolla la interacción cambian, la contribución que se expresa debe adaptarse (o incluso cesar, si deja de estar justificado su desarrollo o si otro participante le arrebata la palabra).

De esta forma, se requiere del Gestor de Toma de Turno la capacidad de conjeturar sobre qué participante recae la posesión de la palabra en cada instante (quién es el hablante), y de actualizar dicha conjetura a medida que cambian las circunstancias de la interacción. Esta estimación se realiza desde una perspectiva de actividad combinada, y en ella sólo serán considerados el estado de los turnos de los distintos participantes y el de los candidatos a tomarla (tal y como describe el sistema de toma de turno de Sacks et al. [149]). El resto de consideraciones de carácter soliciolingüístico que pudieran llevar al sistema a tomar la palabra cuando no existe sobre él un compromiso por parte del resto de participantes para aceptarlo como hablante (dominancia entre roles, el interés por reencaminar el curso de la interacción, etc.) serán aplicadas durante los procesos de decisión de toma de turno, en combinación con el

estado de posesión de la palabra, pero sin serán modeladas dentro de él. Del mismo modo, el sistema aceptará intervenciones primarias de varios participantes simultáneamente, pero siempre estimará cuál de ellos es considerado por el resto de participantes como hablante. Con ello, el estado de posesión de la palabra sólo es revisado, bien como consecuencia de las actualizaciones producidas en el estado de los turnos, o bien por cambios ocurridos en el conjunto de candidatos a tomarla. Al cambiar de manos la posesión de la palabra se desencadenarán procesos de decisión de toma de turno.

El participante en posesión de la palabra es determinado de la aplicación del diagrama de flujo representado en la Figura 25, según el cual, cuando la palabra está vacante (no hay ningún hablante en curso), el primer participante en tomar el turno será considerado el hablante actual (independientemente de la pista de acción, primaria o secundaria, que desarrolle en su contribución). De haber varios participantes interviniendo a la vez (desarrollando contribuciones primarias) se considerará hablante al primero de ellos que comenzó a intervenir. El hablante sólo perderá este título tras haber desistido de su turno (cuando retoma el estado de inactividad), cuando hubiese expresado su fin (al alcanzar el de TRP expresada) o cuando pudiera estimarse que lo ha hecho (al alcanzar el de TRP detectada). En estos casos, la palabra recaerá sobre el primer participante candidato que se encuentre desarrollando turno. Tendrán prioridad los participantes designados frente a los solicitantes, y los que desarrollen turno primario frente a secundario. En caso de no haber ningún candidato, la palabra recaerá sobre el primero que hubiese comenzado a desarrollar una contribución primaria, aunque de no darse ninguno de estos supuestos, la palabra seguiría perteneciendo al hablante en curso.

Ante cualquier cambio en la posesión de la palabra se desarrollarán decisiones de toma de turno que podrían desencadenar procesos de formalización de contribuciones en el Generador de Diálogo (a través del Coordinador de Procesos) con el objetivo de dar al sistema la oportunidad de reformular o interrumpir su contribución en curso, o de evaluar si debe formular una nueva.

6.3.3 Participantes Candidatos a Tomar la Palabra

Los participantes podrán ser apuntados como hablantes siguientes por el hablante en curso cuando éste realice gestos específicos como apuntar con las manos o con la mirada, o utilizando el vocativo. También podrá realizar preguntas específicas, como "*¿y tú qué opinas?*", para motivarle a que tome la palabra. Un participante puede ser también designado como hablante siguiente de forma indirecta, al ser aludido en el desarrollo de un turno previo o cuando el desarrollo del foco actual requiere su participación para poder progresar.

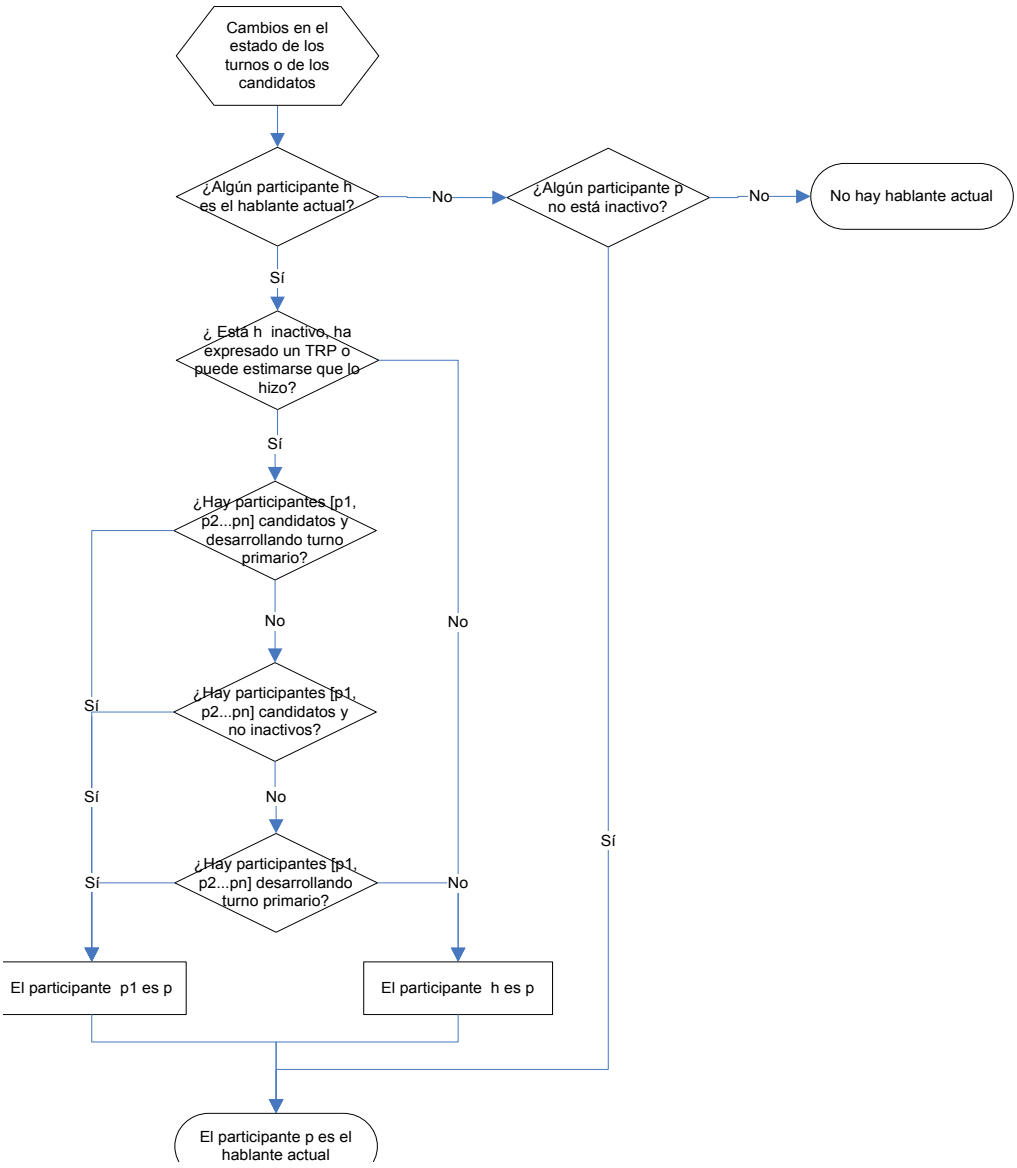


Figura 25: Diagrama de flujo para la estimación del hablante actual

Del mismo modo, los participantes podrán solicitar la palabra durante el desarrollo de la intervención del hablante en curso. Para ello realizarán gestos específicos (como levantar la mano o el dedo), contribuciones habladas (como "*sí, pero…*" o "*disculpe*") o desarrollarán estrategias de ganancia de atención, como comienzo y parada, parada y reinicio, etc. Con todo ello se mostrarán ante el resto de participantes como interlocutores interesados en tomar la palabra, pudiendo repercutir en la reformulación de los turnos que desarrollan actualmente otros participantes para ceder la palabra o retenerla.

En cualquiera de los casos, tales designaciones o solicitudes serán detectadas por el Gestor de Continuidad y por el Gestor de Diálogo durante los procesos de interpretación y generación de diálogo, y serán remitidas al Gestor de Toma de Turno. Éste componente asocia a

cada participante un indicador de su posición frente a turnos futuros, que puede tomar el valor de *designado*, *solicitante* o *nulo* en función de que hayan sido remitidos o no marcadores por parte del Intérprete de Diálogo, durante la interpretación de diálogo de nuevos actos comunicativos (cuando se trata del indicador de un interlocutor), o por el Generador de Diálogo, durante la confirmación de expresión de actos comunicativos (cuando se trata del indicador del propio sistema). También de cuáles sean éstos. La posición de los participantes de cara a turnos futuros se estima, por tanto, a partir de la organización local de la interacción, de su estructura focal y global, de las técnicas aplicadas sobre la continuidad de la contribución y de la evolución de las circunstancias sociolingüísticas en las que se desarrolla. La organización local define transiciones entre estados que pueden ser consideradas cesiones o solicitudes de palabra. También define los participantes de los que depende el desarrollo de cada hilo. De la estructura focal se deduce el hilo sobre el que recae la atención y, por tanto, de qué participantes cabe esperar que tomen la palabra. De la global si el progreso del foco requiere el desarrollo de hilos y, por tanto, determina participantes de los que se puede esperar el desarrollo de contribuciones. Junto a esto, las circunstancias sociolingüísticas pueden aludir a determinados participantes, facilitándoles o requiriendo de ellos la toma de la palabra (si un participante se cae de la silla las circunstancias le hacen protagonista, y todos esperan de él que informe de que se encuentra bien). Por su parte, la gestión de la continuidad permite detectar la aplicación de las técnicas de comienzo y parada, parada y reinicio, etc. que constituyen solicitudes de palabra y otras relacionadas con su cesión. El indicador de posición del participante frente a turnos futuros toma el valor nulo cada vez que se produce un cambio de hablante en la interacción.

De esta forma, la posición de los participantes respecto a turnos futuros es, al igual que el estado de los turnos y la posesión de la palabra, información dinámica que evoluciona con el tiempo a medida que los participantes hacen progresar la interacción y que las circunstancias sociolingüísticas en las que se desarrolla cambian.

6.3.4 Decisión de Toma de Turno

Las decisiones de toma de turno son desencadenadas por los cambios en el estado de los turnos, la posesión de la palabra, los candidatos a tomarla y las metas de la interacción (tanto las discursivas propias como las metas combinadas). Son, por tanto, procesos que parten del Gestor de Toma de Turno y del Gestor de Metas y pueden desencadenar solicitudes de formalización de contribuciones al Generador de Diálogo (tramitadas a través del Coordinador de Procesos.

La decisión de toma de turno permite determinar si existen metas cuyo estado justifica que el sistema tome turno (o lo mantenga) para desarrollarlas (dado el estado actual de los

turnos, de la de posesión de la palabra o los candidatos a tomarla). Para ello se estructura en dos fases: la primera, la calificación de la urgencia con la que cada una de las metas abiertas en la interacción requiere progresar (si debe ser desarrollada de inmediato, si el sistema puede esperar a estar en posesión de la palabra, etc.) y, la segunda, la evaluación de si el estado de posesión de la palabra es favorable o no al sistema (de lo que dependerá que finalmente se formalicen progresos para las metas de cada uno de los niveles de urgencia). Cuando existen metas abiertas para las cuales, dado su nivel de urgencia, el estado de posesión de la palabra sea suficientemente favorable, el Gestor de Toma de Turno solicitará un nuevo proceso de formalización de contribución. El proceso de formalización quedará restringido a las metas marcadas por el Gestor de Toma de Turno como favorables (en base a su nivel de urgencia y al estado de posesión de la palabra) y en él se determinará finalmente si el sistema contribuirá en la interacción (o si continuará haciéndolo); si desarrollará alguna de las metas propuestas; cuál será el orden en que lo hará; y los progresos que supondrán en ellas. Este proceso será desarrollado por el Generador de Diálogo a partir de solicitudes que se cursarán a través del Coordinador de Procesos.

La urgencia de cada meta es calculada a partir de su criticidad, del compromiso alcanzado en su hilo combinado y de la expectativa de beneficio con respecto al coste. La criticidad de la meta es solicitada al Gestor de Metas y hace referencia al interés que tiene el propio sistema en desarrollarla (lo que viene dado por la meta individual que el propio sistema tiene asociada a la meta combinada). Es una función dependiente del estado de interacción y del conocimiento sociolingüístico asociado (según fuese definido por el componente que la introdujo en la interacción). Por su parte, el compromiso es proporcionado por el Estado de Interacción y está relacionado con la calidad de la atención, del interés y de la información que comparten los participantes sobre la meta (con la propia acción combinada con la que se desarrolla la meta en la interacción). Para la calificación de la criticidad y del compromiso, se han identificado experimentalmente cuatro niveles distintos de urgencia: *urgencia baja*, *media*, *alta* y *muy alta*.

Cuando la urgencia de una meta es baja, el sistema no considerará que existan motivos suficientes como para forzar un turno del sistema con el objetivo de hacerla progresar. En definitiva, se formalizará el progreso de la meta si no supone desviar la atención (por ejemplo, a través de modalidades alternativas) o si se trata de una meta secundaria con respecto al hilo actualmente enfocado. En cualquier otro caso, la meta sólo podrá ser desarrollada bajo un estado de posesión de la palabra favorable al sistema.

Si la urgencia es media, el sistema tampoco considerará que existan motivos suficientes para forzar un turno y, al igual que en el caso anterior, el sistema desestimará formalizar progresos para la meta en la interacción (salvo que el estado de posesión de la palabra le sea favorable). No obstante, en los casos en los que desestima formalizar un progreso para la meta, deberá señalizar al resto de participantes su interés por tomar la palabra. Esta señalización se llevará a cabo por medio de la inserción de metas de solicitud de palabra en el Gestor de Metas. Su progreso, al ser de carácter colateral, podrá ser desarrollado en la interacción sin estar en posesión de la palabra (a través de técnicas como levantar la mano o el comienzo y parada) con lo que se auto designa como candidato a tomar la palabra ante el resto de participantes. De esta forma, influirá en el reparto de turnos, favoreciendo el desarrollo de la meta de urgencia media.

Cuando la urgencia de una meta es alta, se considerará oportuno forzar un turno para hacer progresar la meta si fuera preciso, aunque con restricciones. No será necesario estar en posesión de la palabra para formalizar el progreso de la meta, salvo si no se dan unas circunstancias sociololingüísticas favorables (por ejemplo, falta de dominancia del rol del sistema o por cuestiones afectivas o del Auto Modelo). Esto podría suponer la generación de contribuciones del sistema que interfieran en el turno de palabra de otro participante.

Finalmente, la formalización del desarrollo de una meta de nivel de urgencia alto podrá ser realizada en cualquier momento, independientemente de sobre quién recaiga la posesión de la palabra o de las circunstancias sociolingüísticas de la interacción. Se considera aceptable formalizar el progreso de la meta aun constituyendo solapamientos con otras intervenciones en curso o interrupciones.

La urgencia final de una meta vendrá dada por el máximo entre la urgencia de su criticidad y la de su compromiso, y la relación beneficio-coste quedará determinada por los valores de los umbrales que delimitan los distintos niveles. En el caso del compromiso, estos umbrales quedan definidos por el análisis del corpus y, en lo que respecta a la criticidad, son determinados por el componente que introduce la meta.

Para la evaluación de si el estado de posesión de la palabra es o no favorable al sistema, se considerará el estado de los turnos, el de la posesión de la palabra y el de los candidatos a tomarla. Se considerará un estado de posesión de la palabra favorable cuando se den alguna de las siguientes situaciones [Figura 26]: que la palabra esté vacante; que el sistema esté en posesión de la palabra; o que el sistema figure como candidato a tomar la palabra y, a la vez, el turno del hablante haya proyectado un posible TRP, lo haya expresado explícitamente o se pueda conjeturar que lo ha hecho (estados P, D o T del estado del turno del hablante). Cuando la palabra sea favorable al sistema no existirán restricciones en cuanto al conjunto de metas que el

sistema puede desarrollar en la interacción en un nuevo turno (o la continuación de su turno actual). En cualquier otro caso, sólo podrán ser desarrolladas sin consideraciones adicionales las metas muy urgentes, las de plano secundario o las que no requieran desviar la atención. Para el resto de metas, las de urgencia alta sólo podrán progresar sí las circunstancias sociolingüísticas son favorables al sistema y las urgencia media y baja quedarán bloqueadas.

Del mismo modo, si durante la decisión de toma de turno se detecta que el sistema está en posesión de la palabra (pero no existen metas que éste pueda desarrollar en un nuevo turno), el Gestor de Toma de Turno insertará en el Gestor de Metas una meta de cesión de turno (de carácter secundario) que desencadenará la formalización de un turno de rechazo de la palabra (que sería expresado como "No tengo nada que decir", expresiones similares o a través de gestos).

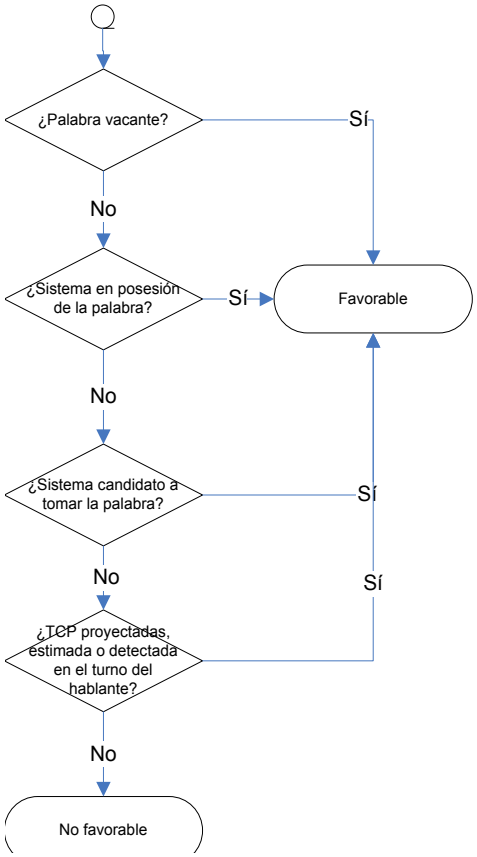


Figura 26: Diagrama de flujo según el cual se determina si el estado de posesión de la palabra es o no favorable al sistema

Todo esto comprende un espacio de decisiones de toma de turno en el que, aunque lo habitual es respetar el turno de palabra del hablante, tienen cabida las designaciones y auto designaciones para optar a la posesión de la palabra, y también la competición por ostentar el

título de hablante. El sistema no se limita a abordar la toma de turno desde una actitud pasiva, primando en la interacción el interés común, pero sin dejar de lado sus propios intereses particulares (aunque esto suponga alterar el reparto de turnos preestablecido). De esta forma, permite soportar los escenarios de toma de turno propuestos por Sacks et al., pero considerando además la influencia de la pista de acción de las metas a desarrollar; las modalidades requeridas para su expresión; y algunas excepciones ocasionadas por las cuestiones circunstanciales que rodean a la interacción. De esta forma, se obtiene una decisión de toma de turno no marcada en la que tienen cabida los fenómenos de solapamiento e interrupción tan naturales a la interacción humana.

6.4 *ESTADO DE INTERACCIÓN*

La propuesta representa la interacción según el Modelo de Hilos, un modelo de diálogo intencional de acción combinada que trata la interacción como el desarrollo de un conjunto de hilos. Según éste modelo del diálogo, el Estado de la Interacción es el componente que mantiene las conjeturas que realiza el sistema sobre la zona común de la interacción, representando la interacción a los niveles local y global. Su estado es actualizado por los componentes Intérprete y Generador de Diálogo durante los procesos de interpretación y generación de diálogo, pero puede ser también consultado por el Gestor de Metas durante las labores de monitorización de la criticidad y el compromiso de las metas.

En lo que respecta a la toma de turno, deben ser consideradas la gestión de versiones que realiza el estado de interacción; la forma en la que se representa el compromiso alcanzado por los participantes sobre el desarrollo de los hilos; la forma en la que se representa la pista de acción sobre la que son desarrollados los hilos; y cómo el estado de interacción permite determinar la designación de determinados participantes como candidatos a tomar la palabra.

6.4.1 Gestión de Versiones

Una toma de turno avanzada requiere la capacidad de proyectar progresos sobre el estado de interacción de forma previa a la realización de las acciones que permiten llevarlos a cabo. El caso más claro es el de la generación, que consta de dos subprocesos realizados por el Generador de Diálogo: la formalización de la contribución y las confirmaciones de expresión de cada uno de los actos comunicativos que la componen. Durante la primera de las fases, el sistema diseña el conjunto de progresos que se realizarán sobre el estado de interacción a medida que el turno transcurra. En la segunda, si no ocurren eventos que hagan interrumpir o

reformular la contribución, van siendo recibidas las confirmaciones de la expresión de los distintos actos comunicativos de la contribución y, con ello, van siendo alcanzados cada uno de los estados de interacción que fueron formalizados inicialmente. Del mismo modo, para proyectar los posibles finales en los turnos de los participantes, el sistema (a través del Adaptador Multimodal y los procesadores de lenguaje natural) debe estimar de forma temprana (antes de haber sido completados) los actos comunicativos que están siendo expresados. Esto también supone anticipar estados de interacción que aún no fueron alcanzados. Por otro lado, de una interpretación incremental pueden surgir (desde el Coordinador de Procesos) solicitudes de reinterpretación de actos comunicativos ya interpretados (lo que obligaría al Intérprete de Diálogo recuperar estados de interacción pasados).

Por todo ello, el conocimiento disponible sobre el estado de interacción es modelado como un histórico de datos en el que se recogen las distintas versiones que se han alcanzado (o están pendientes de ser alcanzadas) de los elementos que contiene. Dado que la unidad de progreso de la interacción es el acto comunicativo, cada versión estará asociada a una instancia de acto comunicativo distinta. De esta forma, cada vez que el Intérprete de Diálogo interprete un nuevo acto comunicativo se alcanza una nueva versión del estado de interacción que se asociará a dicho acto comunicativo. Al ser formalizada una contribución, son proyectadas tantas versiones futuras del estado de interacción como actos comunicativos contenga, y cada una de dichas versiones irá siendo alcanzada con la confirmación de expresión de su acto comunicativo correspondiente. Finalmente, con independencia de la versión que figure como actual en el estado de interacción, ante la solicitud de reinterpretación de un acto comunicativo pasado, es recuperada la versión previa a dicho acto comunicativo (descartándose las versiones posteriores) y sobre ella se construyen los nuevos progresos del estado de interacción.

6.4.2 Gestión del Compromiso

En el Estado de Interacción, los hilos que describen el desarrollo de cada una de las metas que los participantes comparten en la interacción, llevan asociados una estimación del compromiso alcanzado por los participantes sobre su desarrollo. El compromiso de los hilos es actualizado durante los procesos de interpretación y generación de diálogo (desarrollados por el Intérprete y Generador de Diálogo) a medida que sus metas progresan por medio de la interacción y de la "calidad" con la que lo hacen. La propuesta distingue tres aspectos distintos del compromiso: la *atención*, el *interés* y la *información*. Cada uno de ellos consiste en lo siguiente:

- La atención representa el grado en el que en los distintos participantes saben el hilo que se está desarrollando en la interacción en un determinado momento. Es decir, si todos ellos coinciden en aquello de lo creen estar hablando. En la medida en la que el sistema detecte que las contribuciones desarrolladas por el resto de los participantes no encajan en los hilos que él cree que se están desarrollando, se producirán caídas en este aspecto del compromiso.
- El interés representa la relación existente entre un determinado hilo y el hilo enfocado. Los hilos mantienen entre ellos relaciones jerárquicas en la estructura intencional. Cuando un hilo es ancestro del hilo enfocado su interés aumentará, y lo hará en función de su proximidad en la estructura intencional. El nivel de información del hilo enfocado también será modificado por el desarrollo de acciones comunicativas específicas, como los gestos de interés o de desinterés, o expresiones explicitas como "*déjalo*".
- El nivel de información denota la completitud y corrección de la información compartida sobre un determinado hilo. Cuando el sistema carezca de información relevante para el desarrollo del hilo, o detecte inconsistencias en ella, el nivel de dicho aspecto del compromiso se resentirá.

Las variaciones producidas en el compromiso alcanzado sobre los hilos a medida que transcurre la interacción (monitorizadas por el Gestor de Metas) pueden desencadenar nuevos procesos de decisión de toma de turno (desarrollados por el Gestor de Toma de Turno). Estos, a su vez, pueden motivar la generación de nuevas contribuciones del sistema, o la reformulación de su contribución en curso. El Ejemplo 22 muestra cómo las variaciones del compromiso en la información sobre el hilo desarrollado desencadenan constantes reformulaciones de la contribución en curso del sistema. En t=8 y t=9 la realimentación obtenida del interlocutor produce caídas en su valor, lo que lleva al sistema a revisar la contribución con el fin de aumentar el nivel de información que contiene (favoreciendo así la recuperación del compromiso en este aspecto). Finalmente, cuando la realimentación del interlocutor evidencia que el compromiso ha sido restablecido (en t=19), el sistema puede volver a restituir un nivel de información más bajo.

En determinadas situaciones, la reparación o el refuerzo del compromiso pueden requerir la inserción de nuevas metas interactivas propias del sistema. En estos casos, el Estado de Interacción insertará en el Gestor de Metas las metas necesarias para que tales estrategias puedan ser desarrolladas en la interacción en forma de nuevos hilos. El momento en el que estos hilos serán desarrollados en la interacción será determinado por el estado de la toma de turno y por el de la propia meta, durante los procesos de decisión de toma de turno, y la forma en que

se ejecutará su desarrollo será definida en los procesos de formalización (al igual que ocurre con el resto de los hilos de la interacción).

6.4.3 Representación de Pistas de Acción

La estimación del carácter colateral o primario de los turnos activos de los participantes es primordial para una correcta estimación del estado de posesión de la palabra y para hacer posible que el sistema pueda elaborar una decisión de toma de turno natural a la *interacción humana*. La estimación de este carácter pasa por conocer la relación primaria o secundaria que existe entre el hilo que actualmente desarrollan las contribuciones de cada uno de los turnos y el hilo que es considerado foco de la interacción (según la *estructura focal*). Por ello, el Estado de Interacción incluye información sobre la *pista de acción* en que se desarrolla cada uno de los hilos abiertos.

Junto al Estado de Interacción, el Gestor de Diálogo representa la información relativa al dominio para el que se define la interacción. Esta base de conocimiento describe los distintos hilos que podrán tener lugar en la interacción; sus posibles desarrollos; las tareas que corresponde a cada participante realizar en cada uno de sus posibles estados; las transiciones que pueden producirse entre unos estados y otros; y los actos comunicativos que producirán los participantes en la interacción al ejecutar tales transiciones. Del mismo modo, se describen las condiciones bajo las cuales serán abiertas nuevas instancias de los hilos en la interacción (en qué estados de qué hilos tiene sentido que se produzcan, con qué participantes y qué contexto, etc.) y las condiciones iniciales bajo las que serán abiertas (cuáles serán el estado inicial, sus participantes y el contexto de partida). Estos elementos de información se denominan *aperturas*. En la Estructura Intencional del Estado de Interacción se describe la relación jerárquica existente entre las instancias de hilo abiertas en la interacción. Cada nueva instancia de hilo abierta será ajustada sobre alguna de las instancias de hilo ya presentes en el Estado de Interacción. De esta forma, cualquier instancia de hilo tendrá un padre (que será nulo para el *hilo base* de la interacción) y un conjunto de posibles hilos hijo.

Las aperturas descritas en el Dominio de Interacción incluyen información sobre la relación primaria o secundaria que mantendrán las nuevas instancias de hilo que se abran según ella con respecto a la instancia de hilo sobre la que sean ajustadas como hijas (un indicador que podrá tomar como valores: *primaria* o *secundaria*). Junto a ello, cada una de las instancias de hilo contenidas en el estado de interacción estará asociada a una determinada *pista de acción*. Las pistas de acción son elementos del Estado de Interacción también estructuradas jerárquicamente. De esta forma, cada pista de acción tendrá una pista de acción padre (salvo la

pista de acción raíz), y podrá tener pistas de acción hijas. Cada pista de acción será considerada primaria con respeto a cualquiera de sus descendientes, y será secundaria con respecto a sus ancestros. A cada instancia de hilo se le asocia una pista de acción en el momento de su apertura. Para ello, si la apertura según la cual se abrió declaraba una relación primaria con su instancia de hilo padre, dicha instancia de hilo tomará como pista de acción la misma pista de su hilo padre. Si, por el contrario, la relación declarada es secundaria, la instancia de hilo se asociará a una nueva pista de acción creada específicamente para este hilo, y que guardará una relación de hija con respecto a la pista de acción de la instancia de hilo padre.

De esta forma, será posible determinar el carácter primario o secundario de cada uno de los turnos de la interacción en función de la relación primaria o secundaria que exista entre las pistas de acción de la instancia del hilo que desarrolla actualmente y la de la instancia de hilo considerada foco de la interacción.

6.4.4 Influencia del Estado de Interacción en los Candidatos a Tomar la Palabra

Aunque algunos de los marcadores implicados en la designación de candidatos a tomar la palabra y su solicitud son reconocidos por el Gestor de Continuidad, el Adaptador Multimodal y los Intérpretes de Lenguaje Natural, es durante la interpretación de diálogo cuando se identifican los progresos que estos marcadores suponen en la interacción y cuando, por tanto, pueden interpretarse como tales estas designaciones o solicitudes de palabra. Por ejemplo, dependiendo del estado de interacción sobre el que se exprese, un vocativo podrá o no ser interpretado como una designación a candidato a tomar la palabra del participante (puede designar como candidato al participante invocado o, simplemente reclamar su atención). Lo mismo ocurre con el gesto de levantar un dedo (podría ser una solicitud de palabra o un puntero). En definitiva, la interpretación de este tipo de marcadores es dependiente tanto de la formalización del dominio de interacción (dónde quedan definidos las aperturas, los hilos y las transiciones entre sus estados, con los marcadores y actos comunicativos que las desencadenan) y del propio estado en el que se encuentre la interacción (de lo que depende interpretar correctamente lo que significa cada marcador y acto comunicativo en el instante en el que se produce). Cuando el intérprete de diálogo identifica la ocurrencia de una designación de candidato o una solicitud de palabra, éste solicita la actualización del Gestor de Toma de Turno en lo que a los candidatos a tomar la palabra se refiere. Esto podría suponer cambios en el estado de posesión de la palabra y desencadenar incluso nuevas decisiones de toma de turno del sistema (o su cese).

Al igual que es durante la interpretación de diálogo cuando estos marcadores cobran sentido como designaciones y solicitudes durante los turnos de los interlocutores, las confirmaciones de expresión de actos comunicativos realizan la misma tarea para el caso de los turnos del sistema. De esta forma, las solicitudes y designaciones realizadas por el propio sistema reciben el mismo tratamiento que las realizadas por el resto de participantes, favoreciendo la coherencia en la representación que todos los participantes mantienen de la información relativa a los candidatos de la interacción.

Del mismo modo, existen situaciones en las que los participantes no requieren ser designados explícitamente como candidatos a tomar la palabra. En ocasiones la necesidad de que alguno de ellos participe en la interacción para hacerla progresar lo designa automáticamente (por ejemplo, tras una pregunta generalmente está justificado responder, tras un ofrecimiento aceptarlo o rechazarlo, etc.). Por ello, los participantes de los que dependa el progreso del hilo actualmente enfocado en el Estado de Interacción se considerarán también candidatos designados a tomar la palabra. Para reconocerlos, se consultará el estado actual en el que se encuentra el hilo enfocado, se analizarán las posibles transiciones que define hacia otros estados, y se identificará los participantes a los que se asocian.

6.5 *INTERPRETE DE DIÁLOGO*

El Intérprete de Diálogo es el componente que mantiene las conjeturas sobre el estado en el que se encuentran las intenciones particulares de cada uno de interlocutores del sistema en la interacción e interpreta los movimientos que éstos producen con sus contribuciones sobre el Estado de Interacción y resto de conocimiento sociolingüístico asociado. Esto ocurre durante los procesos de interpretación y los de reinterpretación de diálogo.

6.5.1 Interpretación de Diálogo

El Intérprete de Diálogo ofrece el servicio de interpretación de diálogo al Gestor de Continuidad. Sus solicitudes se desencadenan como consecuencia de la detección, en el Adaptador Multimodal y los procesadores de lenguaje natural, de nuevos actos comunicativos en la contribución en curso de alguno de los interlocutores del sistema. Las solicitudes de interpretación son tramitadas a través del Coordinador de Procesos para garantizar que, durante su ejecución, se dispone de un acceso exclusivo a los recursos compartidos que consulta y actualiza [Figura 27]. De entre estos recursos destacan el Estado de Interacción y el Gestor de

Toma de Turnos, aunque también se encuentran en este grupo todos los modelos de conocimiento sociolingüístico asociado (Modelos de Sesión, Situación, etc.).

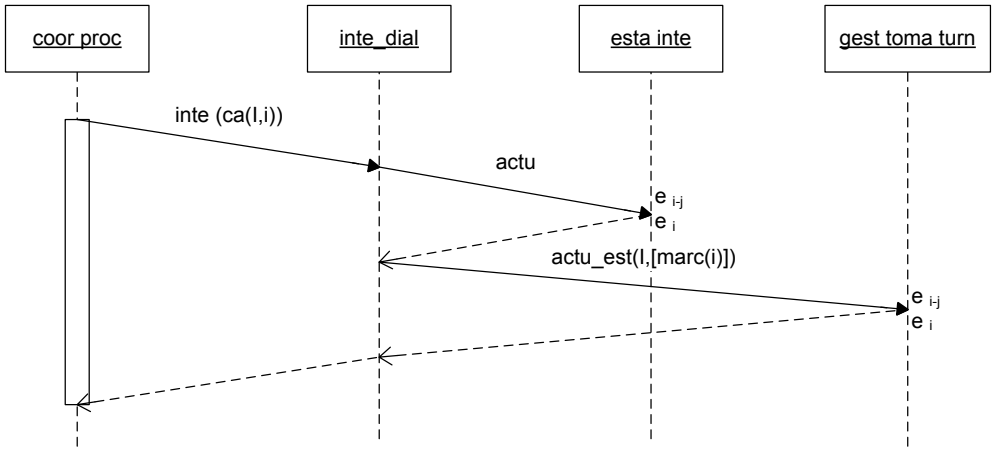


Figura 27: Diagrama de secuencia del proceso de interpretación de diálogo

La interpretación de diálogo de la contribución de un participante no se realiza de forma continua a lo largo del turno de todo el participante, sino a través de subprocesos discretos que se ejecutan puntualmente cada vez que han sido obtenidas nuevas secuencias de actos comunicativos. Durante cada uno de estos procesos son analizados los progresos que suponen los actos comunicativos recibidos sobre el Estado de Interacción. El Estado de Interacción consta de una estructura focal, la cual representa el orden en que han sido desarrolladas las diferentes instancias de hilo en la interacción; de una estructura intencional, que representa las relaciones de dependencia existentes entre ellas; y el estado de desarrollo en el que se encuentran cada una de dichas instancias de hilo. Por su parte, el Dominio de Interacción determina las posibles transiciones hacia otros estados que pueden ser ejecutadas sobre los estados actuales de dichas instancias y de qué actos comunicativos. También determina las aperturas de las nuevas instancias que pueden ser abiertas sobre alguna de las instancias de hilo de la estructura intencional en los estados en los que se encuentran y según el reparto de roles definidos.

Para cada nuevo acto comunicativo recibido, el Intérprete de Diálogo analiza si es posible hacer transitar a algún otro estado alguna de las instancias de hilo abiertas y si es posible ejecutar sobre ellas la apertura de nuevas instancias de hilo (según las estrategias descritas por Calle [21]). Cuando es posible realizar alguno de los progresos descritos sobre el Estado de Interacción, el Intérprete de Diálogo creará una nueva versión del Estado de Interacción asociada a dicho acto comunicativo, y dicha versión pasará a convertirse en la versión actual. En ocasiones no es posible ejecutar ningún progreso sobre el Estado de Interacción para un acto

comunicativo, de forma que no se asociará ninguna nueva versión a dicho acto comunicativo. En otras, como consecuencia de la ambigüedad del acto comunicativo recibido, existen posibles progresos alternativos y, en estos casos, el intérprete de diálogo deberá crear tantas versiones alternativas del Estado de Interacción como sean precisas para representar cada uno de los posibles estados finales. En definitiva, como consecuencia de una posible ambigüedad de la interpretación, pueden darse en el Estado de Interacción interpretaciones alternativas de una misma secuencia de actos comunicativos. El Intérprete de Diálogo deberá estimar cuál de ellas es la interpretación más probable y así marcarla en el Estado de Interacción como versión actual.

La interpretación más probable será aquella sobre la que mejor sea posible interpretar el resto de los actos comunicativos de la contribución. Con cada nuevo acto comunicativo recibido el intérprete de diálogo identificará para cuáles de las versiones alternativas existen posibles progresos desde el estado actual y cuál de ellos se produce sobre una instancia de hilo más cercana al foco. En ocasiones, el Intérprete de Diálogo no contará con elementos de juicio suficientes como para identificar la interpretación más adecuada y, en dichos casos, podrá incluir metas discursivas propias del sistema (a través del Gestor de Metas) para resolver la ambigüedad [Ejemplo 23].

Interlocutor – Avísame de una reunión a las doce y de la comida a las dos. Bueno, mejor a la una.

Sistema – ¿A la una la reunión o la comida?

Ejemplo 23: Resolución de la ambigüedad en la interpretación por inserción de nuevas metas discursivas propias del sistema

Por otro lado, como consecuencia de la coordinación de procesos, la interpretación de diálogo de una secuencia de actos comunicativos no tiene por qué ser realizada sobre la interpretación de diálogo de la secuencia de actos comunicativos previa de la misma contribución, sino que puede ser realizada sobre la interpretación de diálogo de una secuencia de actos comunicativos de otra contribución distinta (de otro participante que pudiera haber tomado turno simultáneamente), o sobre la confirmación de la expresión de un acto comunicativo del propio sistema. De esta forma, es posible procesar simultáneamente las contribuciones que distintos participantes (interlocutores o el propio sistema) pudieran estar realizando a la vez. En la mayoría de los casos se tratará de realimentación simultánea ofrecida a un interlocutor por el resto de participantes durante su intervención. En otros, podría tratarse de contribuciones más complejas que desarrollasen ambas instancias primarias de hilos con respecto a la instancia de hilo enfocada (cuando ocurren luchas por la posesión de la palabra, o son desarrolladas por distintos participantes intervenciones colaborativas).

Del mismo modo, durante el desarrollo de las confirmaciones de expresión de los actos comunicativos de la contribución del sistema, pueden ser interpretados los actos comunicativos de las contribuciones de otros participantes. La presencia de estas contribuciones de otros participantes puede afectar al desarrollo de la contribución en curso del sistema (por ejemplo, en los casos de realimentación simultánea), por lo que la interpretación de diálogo de nuevos actos comunicativos desencadenará la cancelación de la contribución en curso del sistema (como se detalló en el Apartado 6.6.3). Si las razones que llevaron al sistema a formalizar la contribución siguen estando vigentes, el sistema reformulará de nuevo su contribución (y su continuidad será asegurada por el Gestor de Continuidad), aunque es posible que, como consecuencia de la interpretación de diálogo de los nuevos actos comunicativos, el sistema se vea obligado a modificar el desarrollo de su turno (desencadenando reformulaciones cuando el Gestor de Continuidad pueda garantizar la continuidad, o rectificaciones en caso contrario) o a interrumpirlo.

Al igual que ocurre con el Estado de Interacción, la interpretación de diálogo también podría actualizar la información contenida en los modelos de conocimiento sociolingüístico asociado (por ejemplo, la relativa a la sesión) y en el Gestor de la Toma de Turno (según se describió en el Apartado 6.3).

6.5.2 Reinterpretación

En determinadas situaciones, las solicitudes de interpretación de diálogo hacen referencia a actos comunicativos que habían sido previamente interpretados. Esto ocurre cuando, en el Gestor de Continuidad, hay disponible una interpretación de lenguaje natural más precisa de los fragmentos de contribución que dieron lugar a dichos actos comunicativos. O también cuando se identifican errores de interpretación previos. En ambos casos, la interpretación de diálogo de dichos actos comunicativos estará precedida de la restauración del conocimiento sobre el estado de interacción, el estado de la toma de turno y, en ocasiones, el del resto del conocimiento sociolingüístico asociado a la versión previa a la de la anterior interpretación [Figura 28].

Para ello, al ser solicitada la reinterpretación de diálogo de un acto comunicativo, el Intérprete de Diálogo restaura como versión actual del Estado de Interacción la versión que precedía a la interpretación de dicho acto comunicativo, y elimina todas las versiones posteriores a su interpretación (y que con la reinterpretación perderán validez). Una vez restaurada la versión del Estado de Interacción, el Intérprete de Diálogo procederá a realizar una nueva interpretación del mismo, considerando en esta ocasión la información revisada que le

proporciona el Gestor de Continuidad. Los actos comunicativos consisten en una etiqueta que identifica la acción y en un conjunto de parámetros que la completan y contextualizan (sobre el sujeto de la acción, su objeto y otra serie de complementos). La revisión de un acto comunicativo podría incluir nuevos parámetros, valores más precisos en los ya identificados (o la rectificación de errores) y también la actualización de la acción tipificada en el acto comunicativo [191]. La nueva interpretación del acto comunicativo dará lugar a una nueva versión del Estado de Interacción (o varias cuando exista ambigüedad), y su versión actual será actualizada. Al igual que ocurre con el Estado de Interacción, el conocimiento relacionado con el estado de los turnos, de posesión de la palabra y el de los candidatos a tomarla (representados en el Gestor de Toma de Turno) deberá ser también restaurado y posteriormente rectificado. Adicionalmente, deberán ser consideradas las mismas acciones para otros modelos externos a la propuesta, como los son el Modelo de Sesión, de Situación, o el resto de modelos de conocimiento sociolingüístico asociado.

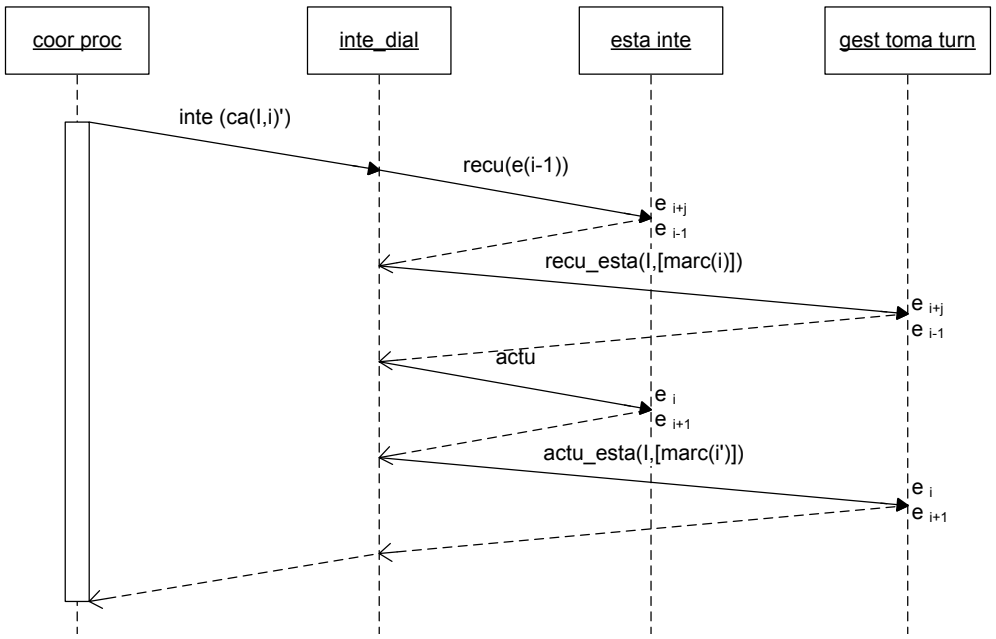


Figura 28: Diagrama de secuencia de la reinterpretación de un acto comunicativo.

Por otro lado, la reinterpretación de un acto comunicativo afectará también a los progresos que fueron realizados, posteriormente a su interpretación inicial, sobre el Estado de Interacción y que están asociados a la interpretación o generación de los actos comunicativos que ocurrieron en la interacción después.

La interpretación de cada acto comunicativo posterior, del mismo u otro interlocutor, habría hecho progresar el Estado de Interacción desde una versión inicial hasta otra final, construida sobre los progresos previos de otros actos comunicativos. Del mismo modo, las

contribuciones formalizadas por el propio sistema habían dado lugar a versiones futuras del Estado de Interacción que podrían incluso haber sido alcanzadas en la medida en la que el transcurso del turno pudiera haber confirmado la expresión de sus actos comunicativos. En definitiva, sobre la versión alcanzada en el Estado de Interacción por la interpretación del un acto comunicativo del que se requiere la reinterpretación, podría existir una secuencia de versiones posteriores que deberán ser revisadas del mismo modo. Por ello, durante la reinterpretación de diálogo de un acto comunicativo, se identifica además la secuencia de actos comunicativos que produjeron progresos posteriores en el Estado de Interacción (tanto de los interlocutores como del propio sistema). Tras la reinterpretación del acto comunicativo, se procederá a revisar la secuencia de actos comunicativos posteriores (en el mismo orden en que ocurrieron).

Si se trata de actos comunicativos de otros participantes, también serán reinterpretados del mismo modo. Si, por el contrario, se trata de actos comunicativos que fueron generados por el propio sistema, deberá evaluarse si su generación, dado el nuevo Estado de Interacción alcanzado, sigue teniendo sentido. En definitiva, se trata de solicitar una formalización de contribución de sistema al Generador de Diálogo y evaluar si el conjunto de actos comunicativos resultantes sería el mismo. En este caso, no se trata de formalizar una futura contribución del sistema, sino de estimar el Estado de Interacción al que habrían llevado los actos comunicativos expresados por el sistema por error y de ejecutar las acciones oportunas para corregirlo. Cuando la contribución formalizada no se corresponde con los actos comunicativos que generó el sistema, deberá analizarse si los actos comunicativos generados erróneamente por el sistema constituyen progresos posibles sobre el Estado de Interacción o si, por el contrario, esto no es así. En el primero de los casos, serán confirmados los progresos que los actos comunicativos del sistema producen sobre el estado de interacción (ya que el sistema los expresó) y, en el segundo, se insertarán en el Gestor de Metas metas discursivas propias del sistema que tendrán como objetivo desarrollar hilos de disculpa. Si estos actos comunicativos desarrollaban hilos que con la rectificación del Estado de Interacción ya no existen, se insertarán metas orientadas a ejecutar su cancelación ("*Ah, entonces nada*"). Por último, cuando se identifique que el sistema ha desarrollado metas cuya urgencia ya no está justificada (cuyos hilos ya no aparecen entre los candidatos marcados por el Gestor de Toma de Turno para ser desarrollados), deberán ser también insertadas metas que permitan restablecer el compromiso de la interacción ("*ay, perdona. Te he interrumpido*").

Todas las metas insertadas por el sistema como consecuencia de los procesos de reinterpretación serán desarrolladas a través de las contribuciones del sistema, al igual que sucede con el resto de las metas discursivas propias del sistema. Su desarrollo se llevará a cabo

cuando sean propuestas por el Gestor de Toma de Turno, siempre que el Generador de Diálogo lo considere oportuno, y en la forma en que éste último determine. En el caso del Ejemplo 12, el sistema habría interpretado que el usuario solicitaba un listado de todos los avisos pendientes, y habría comenzado a generar una contribución para enumerarlos (en t=11). Cuando recibe los fragmentos de contribución de usuario "*para*" y "*mañana*", parte de la contribución habría sido confirmada (alcanzándose versiones parciales de la formalización de la contribución en el Estado de Interacción). La reinterpretación del acto comunicativo del usuario requiere la restauración del Estado de Interacción a una versión previa, y la revisión de los progresos ejecutados posteriormente (en este caso causados por la propia contribución del sistema). De la revisión de la contribución en curso del sistema se detecta un progreso que, dada la nueva interpretación, ahora está fuera de lugar (entre los avisos de mañana no consta ninguna reunión) y el sistema inserta una meta de rectificación encaminada a reparar el compromiso (la que posteriormente dará lugar al desarrollo de la expresión "*¡Ah!, mañana*"). En t=15 la urgencia de las metas de rectificación y listado de los avisos pendientes da lugar a la reformulación de la contribución del sistema (desencadenada por una decisión de toma de turno del Gestor de Toma de Turno y ejecutada por el Generador de Diálogo).

6.6 *GENERADOR DE DIÁLOGO*

La generación de diálogo consta de un proceso previo de formalización de la contribución de sistema y de los sucesivos procesos de confirmación de los actos comunicativos que comprende. Del mismo modo, en ocasiones la formalización de una contribución de sistema es realizada sobre una formalización previa de la contribución que, debido a los cambios ocurridos en la criticidad, en el compromiso de las metas, o en el estado de la gestión de la toma de turno, requiere ser revisada. La responsabilidad de desarrollar los procesos mencionados recae sobre el Generador de Diálogo.

6.6.1 Formalización de Contribuciones

Las solicitudes de formalización son tramitadas por el Coordinador de Procesos y provienen de las decisiones de toma de turno producidas en el Gestor de Toma de Turno como consecuencia de los cambios ocurridos en el estado de los turnos; en la posesión de la palabra; en el conjunto de candidatos a tomarla; y en el estado de las metas. Como respuesta, se obtienen las secuencias de actos comunicativos (que podrían estar vacías) correspondientes a la contribución formalizada o reformulada por el sistema y que serán remitidas al Gestor de Continuidad para que diseñe la forma en que serán expresadas por el canal. El Coordinador de

Procesos garantiza el acceso en exclusiva a los recursos compartidos que aplica (consulta y actualiza) durante la resolución de la solicitud. Este conocimiento es el relativo al Estado de Interacción, Gestor de la Toma de Turnos y el representado en otros modelos externos a esta propuesta (como el Modelo de Sesión) [Figura 29].

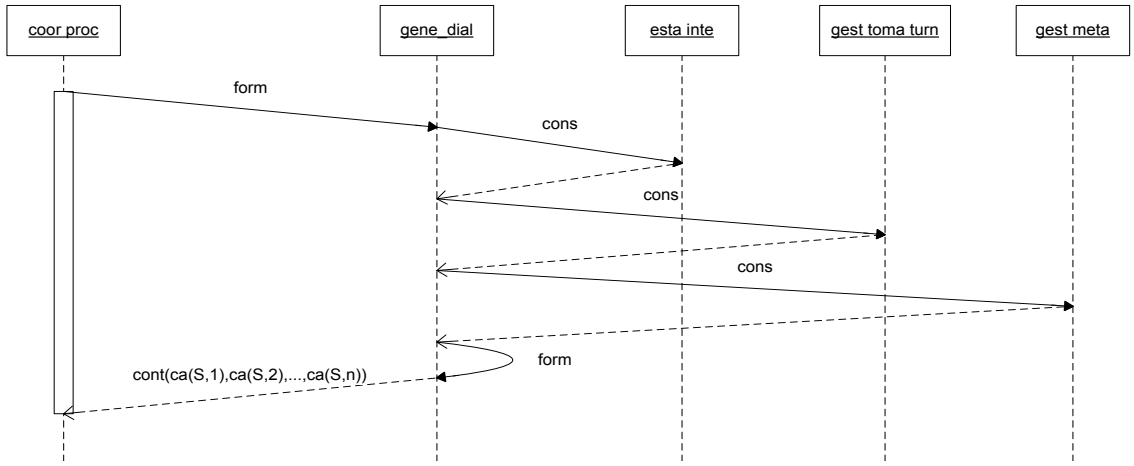


Figura 29: Diagrama de secuencia de la generación de diálogo

Los cambios producidos en el conocimiento representado en el Estado de Interacción y en el resto de modelos de conocimiento (que ocurren como consecuencia de las contribuciones que los participantes realizan, o por la evolución de las circunstancias sociolingüísticas en las que se desarrolla la interacción) alterará la criticidad de las metas discursivas propias del sistema y el compromiso de los hilos combinados, lo que repercutirá en la recalificación de sus niveles de urgencia. Del mismo modo, el desarrollo de los turnos de los participantes irá variando cómo de favorable se presenta el estado de posesión de la palabra a la toma de turno del sistema.

Cuando, durante los procesos de decisión de toma de turno, se identifiquen metas de urgencia suficientemente alta como para ejecutar progresos sobre ellas a través de una contribución del sistema (en función de cómo sea de favorable el estado de posesión de la palabra para el sistema), se desencadenan procesos de formalización de diálogo para construir dichas contribuciones. Las contribuciones se expresan como secuencias de actos comunicativos estructuradas en discursos, cada uno de los cuales se corresponde con la porción de contribución que formaliza progresos para una instancia de hilo concreta. Para formalizar la contribución, el sistema considera las instancias de hilos que desarrollan las metas que el Gestor de Toma de Turno propuso como candidatas para ser desarrolladas (lo que calculó como una función de su nivel de urgencia y de lo favorable que es al sistema el estado de posesión de la palabra). Durante este proceso, el Generador de Diálogo también considera la información recogida en la

estructura foca, la estructura intencional y el estado de desarrollo en el que se encuentra cada una de las instancias de hilo abiertas en la interacción. Del mismo modo, otros tipos de conocimiento sociolingüístico son también considerados (los recogidos en modelos como el de sesión).

Una vez tomada la decisión de contribuir, el Generador de Diálogo revisa, en primer lugar, las instancias de hilo cuyas metas fueron marcadas como candidatas para ser desarrolladas y que se corresponden con los niveles de urgencia alta o muy alta. Estas instancias de hilo serán revisadas de más a menos urgentes y, sobre cada una de ellas, se analizará si es o no posible formalizar progresos asociados al sistema (y cuáles son esos progresos). Tras la formalización de los progresos de dichas instancias de hilo, serán considerados los progresos relativos a las de las metas de urgencia media y baja marcadas como candidatas. En ambos casos, se trata de instancias de hilo cuyo desarrollo no requiere forzar turno, por lo que serán desarrolladas tomando como punto de partida el orden establecido en la estructura focal.

Formalizar el progreso de una instancia de hilo consiste en construir las versiones del Estado de Interacción que recojan las transiciones y efectos producidos sobre la instancia de hilo como consecuencia de la ejecución de las tareas que el rol del sistema lleva asociadas. Para ejecutar las tareas de los estados de las instancias de hilo, el Generador de Diálogo recurrirá al Gestor de Tareas. Los efectos pueden consistir en transiciones a otros estados, la apertura o reapertura de nuevas instancias de hilo, su cierre, la actualización del contexto (o de otros aspectos de la información sociolingüística asociada), la actualización del Gestor de Toma de Turno (con marcadores de designación, solicitud o TRPs), y alteraciones en el orden de la estructura focal. Si entre los efectos de la ejecución una tarea se encuentra la transición a otro estado, los actos comunicativos asociados a dicha transición serán añadidos al discurso de la contribución en la que el sistema formaliza los progresos asociados a dicha instancia de hilo y, del mismo modo, será formalizada una nueva versión del Estado de Interacción que se asociará a dicha versión (estas serán las versiones a ser alcanzadas una vez producidas las confirmaciones de su expresión).

En ocasiones, el desarrollo de una instancia de hilo requiere, para progresar, el desarrollo previo de alguna de sus instancias de hilo hijas (o incluso el desarrollo de alguna de las instancias ancestro, cuando no alcanzaron aún el estado que justifica su apertura). En estos, previamente al desarrollo de la instancia de hilo, serán formalizados los desarrollos de estas instancias hijas o ancestros, aun sin que sus metas hayan sido marcadas como candidatas por el Gestor de Toma de Turno. En otras, el progreso de una instancia de hilo cuya meta hubiese alcanzado una criticidad elevada podría no depender del sistema (para cuyo rol podrían no estar

descritas transiciones desde el estado en el que se encuentra). En dichos casos, el Generador de Diálogo incluirá nuevas metas discursivas propias en el Gestor de Metas con el objetivo de favorecer participaciones del los interlocutores que la hagan progresar ("*entonces, ¿qué opinas de…?*", o metas de relleno, en el caso del hilo base).

Durante la generación podrán alcanzarse estados finales en alguna de las instancias de hilo abiertas en la interacción. Cuando esto sucede y sobre dicha instancia existe una meta discursiva propia del sistema, ésta podrá darse por resuelta y, tal hecho, deberá así ser notificado al Gestor de Metas. Durante la formalización de una contribución serán proyectadas tales notificaciones, aunque su ejecución será realizada durante la confirmación de expresión de los actos comunicativos a los que tal resolución está asociada.

Los actos comunicativos resultantes serán sometidos a un proceso de generación de referencias deícticas que dotarán a la contribución de una mayor naturalidad (para ello se aplicarán otros componentes, como los Modelos de Sesión o Situación). Finalmente, una vez formalizada completamente la contribución del sistema, ésta es enviada al Gestor de Continuidad, para que coordine su expresión a través de modalidades y lenguajes naturales a sus interlocutores.

6.6.2 Confirmación de Síntesis

Durante los procesos de formalización de contribuciones, el Estado de Interacción no se actualiza, proyectándose los progresos futuros que se desencadenarán en él a medida que los actos comunicativos que las componen vayan siendo expresados (pero sin ejecutar dichos progresos). En definitiva, tras la formalización de una contribución, habrán sido creadas las versiones futuras del Estado de Interacción que se alcanzarán tras las expresiones de sus actos comunicativos. Será durante las confirmaciones de dichas expresiones (confirmaciones de síntesis) cuando el Estado de interacción vaya progresando a través de dichas versiones. Así ocurrirá también con los cambios proyectados para el Gestor de Toma de Turno (producidos por la detección de marcadores como las designaciones de candidatos, las solicitudes de palabra y las TRPs) y con el conocimiento referente a otros componentes externos a la propuesta, como lo es el Modelo de Sesión. Durante la ejecución de las confirmaciones de expresión son también confirmadas la resolución de las metas asociadas a las instancias de hilo que alcanzan sus estados finales en el Gestor de Metas.

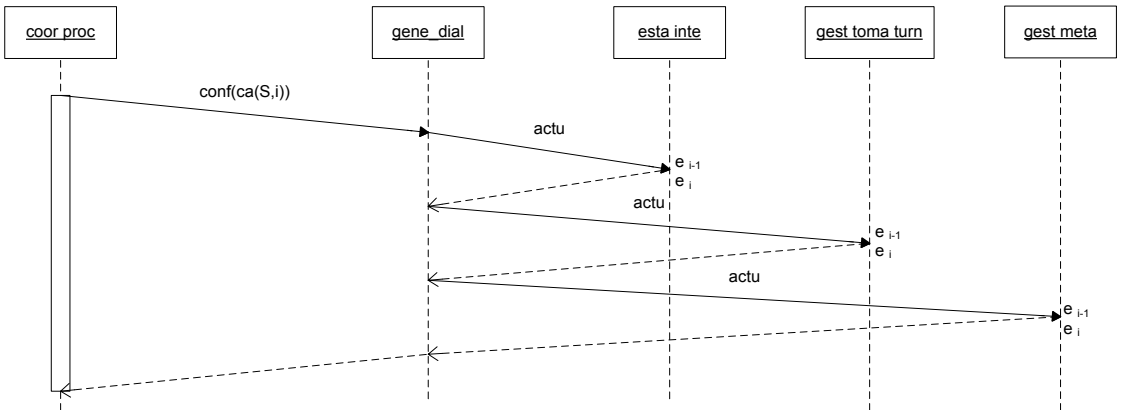


Figura 30: Diagrama de secuencia de confirmación de un acto comunicativo

6.6.3 Reformulación de Contribuciones y Auto Interrupciones

Si durante la producción de la contribución previamente formalizada del sistema cambian las circunstancias que rodean a la interacción, en aspectos tales como la criticidad de las metas discursivas propias del sistema; los hilos combinados del sistema; o el estado de posesión de la palabra, serán desencadenadas nuevas solicitudes de formulación de contribución de sistema que provocarán una reformulación o interrupción de la contribución.

El proceso pasa por la cancelación de la generación de la porción de contribución aún pendiente (realizada por el Coordinador de Procesos), tras lo cual se desarrolla una nueva formalización. Ésta habrá sido desencadenada por un nuevo proceso de decisión de toma de turno del Gestor de Toma de Turno. En la medida en la que las condiciones que desencadenaron la contribución previa se mantengan, la reformulación dará lugar a una contribución similar a la porción de contribución formalizada previamente (cuya expresión quedó pendiente de confirmación). En función de que hayan aparecido nuevas metas; de que haya sido cancelada o resuelta alguna de las existentes; de que su urgencia se haya modificado; de que se hayan producido otros progresos sobre las instancias de hilo que las desarrollan; o de que haya variado el estado de posesión de la palabra o de los candidatos a tomarla, está contribución será distinta y, en resultado, será una reformulación suave, rectificación o auto interrupción en función de lo bien que pueda ser concatenada en el Gestor de Continuidad con la que fue cancelada.

De esta forma, será posible adaptar la contribución en curso del sistema a la realimentación simultánea ofrecida por los interlocutores y a los cambios en las circunstancias sociolingüísticas en las que se desarrolla el turno del sistema. Los eventos ocurridos en la posesión de la palabra son algunos de los escenarios en los que pueden ocurrir. Se gestionarán a través de reformulaciones las siguientes situaciones:

- El sistema comenzó a formalizar su contribución sin estar en posesión de la palabra (restringiendo el conjunto de metas a desarrollar) pero a lo largo del turno pasa a ser el hablante (y por tanto puede tratar en su turno metas de urgencia más baja).
- El sistema formalizó su contribución estando en posesión de la palabra pero durante el desarrollo del turno la pierde (por ejemplo, si otro participante es designado en él y decide tomar turno primario en algún punto que puede ser entendido como TRP).
- Conclusión prematura de su contribución cuando otro participante solicita la palabra y el sistema considera menos prioritario el desarrollo de sus propias metas.
- Interrupción de su contribución cuando interviene simultáneamente a otro participante y la situación no le hace dominante.
- Desarrollo de estrategias de mantenimiento de la palabra (levantar la voz, hablar más lento) cuando otro participante interviene simultáneamente y la situación le hace dominante o la urgencia de sus metas le obliga a intervenir.
- Extensión de su contribución cuando es posible para el sistema desarrollar más metas y ningún otro participante reclama la palabra.

6.7 *GESTOR DE METAS*

Cada acción combinada en la interacción está compuesta por las acciones individuales de cada uno de los participantes. En el caso de la acción correspondiente al sistema para cada una de las metas que se desarrollan en la interacción, se trata de las metas del sistema. Éstas quedan representadas en el Gestor de Metas.

El Gestor de Metas es el componente encargado de recibir las metas discursivas propias del sistema que los distintos componentes de la arquitectura insertan; de recibir sus cancelaciones y resoluciones; y de evaluar cómo de crítico es para el sistema desarrollarlas en cada instante, así como de monitorizar el estado del compromiso alcanzado en el desarrollo de las distintas instancias de hilo del Estado de Interacción. En la medida en la que se produzcan aumentos significativos en los niveles de criticidad de las metas discursivas propias del sistema (o caídas en alguno de los aspectos del compromiso de los hilos combinados), el sistema desencadenará decisiones de toma de turno, en el Gestor de Toma de Turno, que podrán tener

como consecuencia el desarrollo de nuevos procesos de formalización de contribuciones en el Generador de Diálogo. A continuación se describe cada uno de dichos procesos.

6.7.1 Inserción, Cancelación y Resolución de Metas

Cualquiera de los componentes de la arquitectura podrá incluir en cualquier momento nuevas metas discursivas propias de sistema con el objetivo de que sean resueltas a través de la interacción. Por ejemplo, esto sucede para desarrollar estrategias de refuerzo y reparación; cuando el sistema desea resolver ambigüedad en la interpretación o comunicar otro tipo de problemas en ella ("*hay mucho ruido, no le entiendo*"); cuando desea desarrollar solicitudes de toma de turno o turnos de paso; cuando se desea comunicar al interlocutor la ocurrencia de determinados eventos espacio-temporales ("*¡Cuidado!, el suelo está mojado*"), cuando se requiere determinada información del usuario ("*¿A qué hora quiere programar la alarma?*"); etc. Durante la inserción de una meta, el componente que la solicita determina algunos parámetros relacionados con su desarrollo y su resolución. Entre ellos, se encuentran la función de su criticidad y los umbrales según los cuales se fijarán sus niveles de urgencia. Junto a ellos podrá establecerse un contexto inicial; unas relaciones de dependencia con respecto a otras metas de la interacción; y las líneas del contexto que, tras la resolución de la meta, serán consideradas su resultado.

Algunos ejemplos de metas que pueden insertar los distintos componentes del sistema son: solicitar información al usuario, cuando la aportada en la interacción hasta el momento es insuficiente ("*¿A qué hora quiere programar la alarma?*"); avisarle de eventos espacio-temporales ("*¡Cuidado!, el suelo está mojado*"); comunicarle problemas de interpretación ("*hay mucho ruido, no le entiendo*"); etc. Al recibir nuevas metas, el Monitor de Metas las incluye en un espacio de metas individuales genérico (no asociado a sesión).

Cuando una meta es insertada en el Gestor de Metas se solicita al Generador de Diálogo la asociación de alguna instancia de hilo a dicha meta para su desarrollo. El Generador de Diálogo considerará para ello la compatibilidad de la meta con las instancias de hilo ya abiertas (en función de la intención que desarrollen y su contexto). En caso de no ser posible ajustar ninguna de las instancias de hilo ya abiertas a la meta, el Generador de Diálogo le asociará una nueva instancia de hilo compatible con ella. Del mismo modo, la meta pasará a ser considerada en las decisiones de toma de turno desarrolladas por el Gestor de Toma de Turno en igualdad de condiciones a las ya existentes. De esta forma, las variaciones en su criticidad y en el compromiso podrán desencadenar la formalización o reformulación de las contribuciones del sistema.

El componente que insertó una meta discursiva propia en la interacción podrá solicitar su cancelación en cualquier momento al Gestor de Metas. Tal hecho es notificado al Generador de Diálogo, quien analiza si la instancia de hilo de la meta a cancelar ha comenzado ya a progresar en la interacción, en cuyo caso insertará en el Gestor de Metas una nueva meta discursiva propia que permitirá desarrollar en la interacción segmentos orientados a notificar a los interlocutores que dicha meta fue cancelada.

Durante los procesos de generación de diálogo, las distintas instancias de hilo de la interacción irán progresando hasta haber alcanzado estados finales. Cuando esto ocurre, la meta discursiva propia que el sistema asocia al desarrollo de dicha instancia de hilo podrá darse por resuelta. En estas situaciones, el Generador de Diálogo notifica al Gestor de Metas la resolución de la meta quien, a su vez, se lo comunicará junto con el resultado obtenido al componente que la insertó.

6.7.2 Monitorización de Metas Discursivas Propias

El valor de urgencia de una meta viene dado por su función de criticidad, que es determinada por el componente que insertó la meta y describe como la criticidad de la meta evoluciona a medida que el estado de interacción y el conocimiento sociolingüístico cambian. Puede ser dependiente, por ejemplo, del tiempo transcurrido desde su inserción o desde la última vez que la meta tomó el foco; del valor de determinadas líneas de contexto; del estado del compromiso de su hilo o del de otros hilos; etc. Para permitir la actualización de la criticidad de las metas al cambiar las variables de las que depende, el Monitor de Metas podrá suscribirse a notificaciones de cambio de dichas variables en los modelos que las recogen. Para ello, contará con un subcomponente Gestor de Variables.

Por ejemplo, la criticidad de la meta discursiva propia del sistema sobre la propia conversación, haya sido iniciada o no por él, debe depender del tiempo de inactividad desde la última contribución. De esta forma, cuando se producen silencios incómodos, la urgencia de la meta discursiva propia sobre la instancia de hilo base de la interacción se disparará, lo que llevará al sistema a intentar desarrollar dicha instancia de hilo. Dado el proceso de formalización de nuevas contribuciones descrito en el Generador de Diálogo, el desarrollo de la instancia de hilo base dependerá del desarrollo de sus instancias de hilo hijas. Por ello, el Generador de Diálogo tratará, en primer lugar, de desarrollar alguna de sus instancias de hilo hijas, o alguna de sus descendientes. Si el desarrollo de las instancias de hilo de la interacción depende de otros interlocutores, tratará de incitarlos a desarrollar progresos sobre dichos hilos ("*Sobre el aviso que teníamos pendiente, ¿a qué hora desea programarlo?*"). Si no existen

instancias de hilo que puedan ser desarrolladas por el sistema, no existen instancias de hilo que puedan ser desarrolladas por otros interlocutores (o dichos interlocutores no responden a las sugerencias del sistema por hacerlos progresar), el sistema tratará de insertar nuevas metas de relleno con el objetivo único de rellenar los silencios ("*pues, hace bueno*"). Por último, si no es posible hacer progresar la interacción, el sistema la cerrará.

Cuando las variables pertenecen a modelos de conocimiento cuyo estado es versionable y es posible navegar entre ramas alternativas del árbol o retroceder a versiones anteriores (como es el caso del Estado de Interacción), también serán recibidas notificaciones de cambio de dichas variables. Esto ocurrirá como consecuencia de cambios de versión, producidos tras procesos como el de reinterpretación o el de confirmación de actos comunicativos.

A medida que las distintas variables manejadas por los distintos modelos de conocimiento actualizan su valor (bien por efecto de contribuciones de otros participantes o del propio sistema, por cambios en la situación, etc.), la criticidad de las metas discursivas propias del sistema aumentará o disminuirá. Al cambiar el valor de criticidad de una meta (bien al aumentar o al reducirse) se desencadenarán nuevas solicitudes de formalización de contribución (a ser desarrolladas por el Gestor de Toma de Turno). Ante estas solicitudes, dicho componente podrá decidir en función del estado de las metas, de los turnos, de la posesión de la palabra y de los candidatos a tomarla, si propone o no el desarrollo de tales metas para un nuevo proceso de formalización de contribución del sistema (que desarrollará el Generador de Diálogo).

En ocasiones, la interacción finaliza sin que todas las metas discursivas del sistema hayan podido ser resueltas. En ese caso, el Monitor de Metas reasignará la gestión de dichas metas al Auto Modelo quién, en la medida en que su criticidad vuelva a dispararse, procederá a asignarla a alguna otra sesión abierta sobre la que pueda ser desarrollada, o iniciará una nueva sesión para resolverla ("*Disculpe que le interrumpa, pero tiene una cita programada.*").

6.7.3 Monitorización de Hilos Combinados

Cuando el compromiso de alguna de las instancias de hilo es actualizada durante los procesos de interpretación (incluida la reinterpretación) o de confirmación de actos comunicativos (según los progresos diseñados para las instancias de hilo durante las formalizaciones de contribuciones), el Gestor de Metas consultará el nivel de compromiso alcanzado en sus distintas componentes (atención, interés o información) y, cuando se den en ellas caídas pronunciadas, según determinen los niveles de urgencia que define para dicho hilo el Dominio de Interacción, se procederá a ejecutar acciones encaminadas a restablecer su nivel.

Estas acciones consistirán en la inserción de metas discursivas propias del sistema. Con ellas podrá desarrollar pausas o actos nulos para reforzar la atención, el compromiso o el interés (cuando su caía no es especialmente brusca). Para un refuerzo mayor, y cuando la caída afecta a la información o a la atención, puede desarrollar hilos de redundancia o aumentar el nivel de información que acompaña a los actos comunicativos de sus contribuciones. En esta misma línea, también podrá enumerar la línea de ancestros del hilo enfocado, lo que mejorará en especial la atención. Finalmente, cuando la caída de alguno de estos aspectos es brusca, el sistema recurrirá a interrupciones directas. Algunos ejemplos son los siguientes:

- Refuerzo de atención: *"Me he perdido. ¿De qué estamos hablando?"*.
- Refuerzo de interés: *"¿Quiere seguir programando el aviso?"*
- Refuerzo de información: *"No me refería a la hora de la reunión, sino a la de la comida"*.

Capítulo 7 **EVALUACIÓN**

Este apartado describe la evaluación a la que ha sido sometida la propuesta de Sistema de Interacción Natural de esta tesis doctoral. La evaluación se centra en las habilidades específicas de una toma de turno avanzada, dejando de lado el resto de habilidades que comprende el tratamiento de la *interacción natural* (reconocimiento y síntesis de lenguajes naturales; procesamiento de lenguaje natural; adaptación multimodal; gestión de diálogo a los niveles local y global; gestión del conocimiento sobre las circunstancias sociolingüísticas que envuelven a la interacción; y representación de la ontología de conocimiento). Se propuso una metodología de evaluación comparativa de un mismo sistema de interacción bajo diferentes configuraciones de toma de turno: estrategia de toma de turno propuesta; toma de turno por ciclo de interacción; y toma de turno humana. Durante los experimentos fueron recopilados tanto parámetros técnicos del funcionamiento del sistema como valoraciones subjetivas de la naturalidad de la interacción percibida por el usuario. Entre los parámetros técnicos considerados se encuentran: el tiempo de interacción; el porcentaje de tiempo de posesión de la palabra de cada participante; el número y tipo de decisiones de toma de turno; y el número de solapamientos e interrupciones ocurridas en la interacción. En lo que respecta a la evaluación subjetiva de la naturalidad, se valoraron parámetros como la satisfacción subjetiva manifestada por el usuario o su percepción sobre lo dinámica y ordenada de la interacción; la actitud colaborativa del sistema; lo adecuado de la estrategia de toma de turno para la resolución de la tarea propuesta; y lo cómoda que resulta la estrategia interactiva desarrollada.

7.1 DOMINIO DE INTERACCIÓN

Con el objetivo de evaluar la naturalidad de la toma de turno del sistema, se buscaron dominios de interacción en los que se pusieran en juego habilidades avanzadas de toma de turno y en los que fuera minimizada la influencia de los problemas asociados a otros niveles de la interacción natural (reconocimiento y síntesis de voz, procesamiento de lenguaje natural, adaptación multimodal, etc.). Del mismo modo, se evitaron dominios en los que la modalidad dominante para el usuario fuese el habla (por los problemas asociados al reconocimiento de voz y el procesamiento de lenguaje natural en situaciones de habla espontánea).

Para la elección de un dominio de interacción adecuado fueron analizados varios de los dominios pertenecientes al ámbito del desarrollo del proyecto SemAnts (TSI-020100-2009-419), que comprendían juegos y tareas colaborativas. De cada uno de estos dominios se recopiló un corpus preliminar, aplicando la técnica persona-persona (dos personas realizando interacciones en escenarios del dominio propuesto, una en el papel de sistema y otra en el de usuario). Durante la adquisición, los participantes no fueron instruidos sobre el guión específico que debían desarrollar, ni sobre el conjunto de expresiones, modalidades o reglas de toma de turno a los que debían restringirse.

Tras el análisis de los corpus adquiridos, el dominio de interacción seleccionado fue el "dominio de dictado" [Figura 32]. Este dominio está caracterizado por dos roles (sistema o jefe y usuario o secretario) y por un objetivo claro y comprometido por ambos participantes (que el texto que dicta el jefe sea copiado íntegramente y sin errores por el secretario). El dominio de dictado se perfila como un dominio muy adecuado para la evaluación de la toma de turno. En este dominio, toda la interacción se desarrolla en torno a una contribución primaria del sistema que se desarrolla desde el principio hasta el final de la interacción (aquella con la que dicta al usuario), y en la que el sistema va notificando correcciones al texto copiado. Este dominio se caracteriza por la sencillez de la tarea a resolver, el reducido conjunto de tipos de segmentos que los participantes desarrollan para ello, y un número limitado de acciones comunicativas que, de forma natural, se desarrollan en él. Junto a esto, en sus contribuciones se observan gran cantidad de reformulaciones, interrupciones y auto interrupciones. Es por ello que, este dominio, resulta muy adecuado para la evaluación de una toma de turno avanzada.

Desde el punto de vista de la reformulación y la gestión de la continuidad, destaca el hecho de que la contribución del sistema se ve claramente afectada por los cambios producidos en las circunstancias que envuelven en la interacción (el ritmo al que copia el usuario, la ocurrencia de errores en el texto copiado, etc.).Del mismo modo, en la interacción se desarrollan

tanto contribuciones primarias como secundarias (especialmente las realizadas por el usuario). De entre las de carácter primario destacan las preguntas sobre ortografía, o las solicitudes de repetición. Como contribuciones secundarias caben ser destacadas las solicitudes de palabra, las manifestaciones de inquietud o inseguridad, o la propia actualización del texto copiado (que ofrece una constante realimentación a la contribución que desarrolla el sistema).

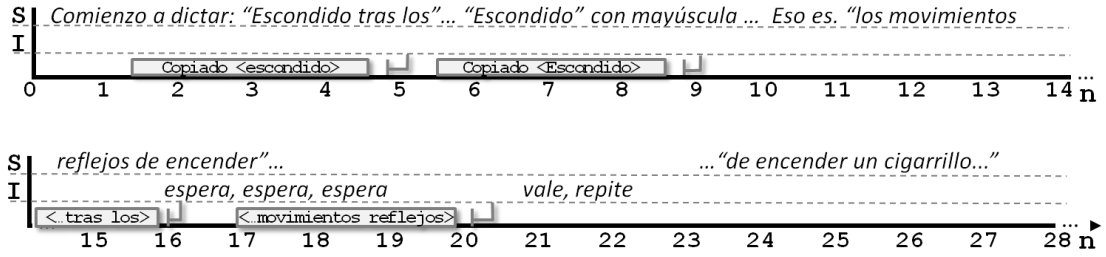


Figura 31: Fragmento de interacción del corpus de evaluación del domino de dictado

En lo que respecta a la toma de turno, cabe destacar que ambos participantes contribuyen simultáneamente en la interacción (el sistema dictando y el usuario copiando) y que, en los casos de requerir ambos participantes tomar la palabra a la vez, pueden darse situaciones diversas: El usuario podrá decidir si solicita la palabra para intervenir o si interrumpe al sistema. El sistema, por su parte, si la cede o no (en función de que haya sido alcanzado un punto adecuado para ello en el desarrollo del dictado).

Junto con todo esto, el dominio de dictado permite el desarrollo de la gestión de metas, coexistiendo en el estado de interacción la meta que constituye el propio dictado con las de corrección de errores ortográficos o tipográficos que incluye el sistema, o las de solicitud de información ortográfica, recapitulación o repetición que incluye el usuario (entre otras). El corpus resultante está compuesto de un total de ocho hilos, de entre los que destacan: el hilo de dictado; el hilo de notificación de errores ortográficos o tipográficos en el texto copiado; preguntas al sistema sobre dudas ortográficas; solicitudes de repeticiones parciales o completas del texto copiado; solicitudes de parada y continuación; deletreado de palabras; y otros hilos de refuerzo cuando el compromiso decrece [Tabla 9].

7.2 *USUARIOS Y SESIONES*

Los experimentos de este trabajo fueron realizados en septiembre de 2010 en los laboratorios del Grupo LaBDA de la Universidad Carlos III de Madrid. En ellos participaron un total de 39 personas (28 hombre y 11 mujeres), de los cuales todos eran hablantes de español

(35 hablantes de castellano, tres hablantes de mejicano y uno hablante de colombiano). Sus edades estaban comprendidas entre los 14 y los 62 años (media: 29,98; desviación típica: 11,57) [Tabla 10]. De todos ellos, 33 usaban con frecuencia ordenadores y los otros seis sólo ocasionalmente o nunca. Cinco participantes no tenían estudios regulados; dos de ellos contaban con graduado escolar; dos con título de bachillerato; 12 tenían estudios universitarios; 11 estudios de postgrado; seis se consideraban expertos en las Tecnologías de la Información; y uno experto en Interacción Hombre Máquina. Todos los participantes fueron encontrados a través de anuncios en los tablones de la universidad y todos ellos guardan alguna relación con ella (bien sea estudiantes, personal de servicio, profesores u otro tipo de visitantes).

Durante los experimentos se recopilaron 141 sesiones de interacción con el sistema, cada una de ellas conteniendo una media de seis minutos de diálogo y alcanzándose un total de casi cuatro horas de corpus. De éstas, 78 minutos se correspondían con interacciones según una toma de turno humana, 72 minutos con interacciones desarrolladas bajo la estrategia de toma de turno propuesta y 101 minutos con interacciones desarrolladas por ciclo de interacción. Se obtuvieron una media de 50,05 palabras por interacción (33,66 en la toma de turno avanzada, 78,28 en el ciclo de interacción y 38,2 en la toma de turno humana). Algunos casos de reparaciones, vocalizaciones no verbales, risas y carraspeos, entre otros, fueron anotados como palabras. Todas las sesiones fueron grabadas en video.

Tabla 9: Hilos formalizados en el domino de dictado

Hilo	Descripción
Dictado	El sistema dicta al usuario un texto
Corregir error	El sistema corrige un error ortográfico o tipográfico en el texto copiado por el usuario
Consulta ortográfica	El usuario desea comprobar la ortografía de una determinada palabra
Repetición	El usuario solicita la repetición parcial o total del texto del dictado
Recapitulación	Técnica de refuerzo
Deletreado	El sistema deletrea una palabra (por su propia iniciativa o por solicitud del usuario)
Parada/continuación	El usuario solicita una pausa en el dictado y lo retomará posteriormente
Disculpa	El usuario se disculpa por los errores

7.3 *DISEÑO DE LOS EXPERIMENTOS*

Ambos participante, usuario y sistema, tienen como objetivo común completar un dictado durante cada interacción. Dado que la tarea sólo se considera resuelta cuando el texto copiado por el usuario es exactamente igual al dictado por el sistema, los participantes deberán cuidar aspectos como la ortografía, los signos de puntuación y el uso de mayúsculas. Ni el usuario ni el sistema tienen la capacidad de resolver la tarea por si solos, puesto que el sistema

no puede copiar el texto y el usuario no conoce de antemano qué texto debe copiar. Este hecho garantiza un interés común por parte de ambos participantes en el desarrollo de la interacción y su colaboración para resolver la tarea. Durante el dictado el sistema puede parar de dictar para ayudar al usuario a corregir sus errores. También puede adaptar dinámicamente el ritmo de dictado a las necesidades del usuario. En lo que respecta al usuario, se espera de él que se comporte como un humano, preguntando en todo momento sus dudas ortográficas, solicitando paradas y continuaciones, repeticiones, etc.

Tabla 10. Caracterización de sujetos de evaluación

	Hombres	**Mujeres**
Sexo	28	11

	< 25	**25-34**	**35-44**	**>=45**
Edad	9	16	8	6

	Sin estudios	**Graduado escolar**	**Bachillerato**	**Estudios universitarios**	**Estudios de postgrado**	**Experto en TIs**	**Experto en IHM**
Formación	5	2	2	12	11	6	1

	Ocasional / nunca	**Frecuente**
Uso de PC	6	33

Cada usuario debe realizar un total de tres dictados diferentes con el sistema, donde cada uno de ellos será realizado bajo una estrategia de toma de turno diferente. Estas configuraciones son:

A) Toma de turno por ciclo de interacción (los turnos son desarrollados en un orden de participación secuencial y predefinido)

B) Toma de turno avanzada (el sistema desarrolla la estrategia toma de turno propuesta en este trabajo)

C) Mago de Oz (un humano determina qué debe decir el sistema y cómo debe comportarse en cada momento).

Estas configuraciones se presentan a los usuarios en un orden aleatorio, de tal forma que no pueden conocer cuál es la estrategia de toma de turno que desarrolla en sistema en cada momento. Los usuarios tampoco conocen cuál de las configuraciones es la que está siendo sometida a evaluación, ni qué parámetros se consideran para ello. Todos los usuarios son previamente entrenados en la tarea de copiar textos en las mismas condiciones en la que,

posteriormente, se desarrollan los experimentos (con la misma interfaz de usuario y en el mismo entorno). El entrenamiento consiste en una breve explicación tanto de la tarea a realizar como de las herramientas con las que se cuenta para ello. Durante esta fase también se realizan tantos dictados como el usuario necesite para garantizar que se comporta como un experto en la tarea (copiar dictados con el ordenador) durante la interacción. Los textos usados durante los entrenamientos y los dictados son elegidos aleatoriamente, siguiendo una distribución uniforme, de entre un conjunto de textos de similares características (70 textos infantiles de entorno a 200 caracteres cada uno).

Con el objetivo de homogeneizar la forma en la que las distintas configuraciones del sistema se muestran a los usuarios, se incluyen en los experimentos dos nuevos participantes, un líder y un mecanógrafo, cuyos roles son actuar como interfaz entre el usuario y el sistema durante la interacción. Para ello, los participantes (usuario, líder y mecanógrafo) cuentan con un ordenador cada uno y el usuario se sienta frente al líder, mientras que el mecanógrafo permanecerá en un segundo plano [Figura 33]. Durante el dictado, el sistema muestra a cada participante una interfaz gráfica distinta, en función de las tareas específicas que deben realizar [Figura 34]. Para el usuario, muestra un campo de texto en el que copiar el texto dictado. Para el líder, el sistema muestra las expresiones que éste tiene que expresar y la forma en que debe hacerlo (ritmo, entonación, etc.). Finalmente, el mecanógrafo tiene como interfaz un área de texto en el que copiar las expresiones que el usuario realiza. Para agilizar el proceso de adquisición de las expresiones de usuario y complementarlo con expresiones realizadas a través de otras modalidades, la interfaz del mecanógrafo se completa con un conjunto de botones que representarán las expresiones verbales y gestuales más frecuentes del dominio.

Figura 32: Disposición de los participantes del experimento en el entorno de evaluación

Cuando el usuario interactúa con una configuración máquina (A o B), el líder asume la responsabilidad de sintetizar las expresiones que el sistema le muestra a modo de *karaoke*. En este caso, el sistema va escribiendo su contribución sobre la interfaz y el líder va leyéndola al ritmo que el sistema le presenta (marcando también silencios, silencios oralizados, etc.). En este caso, el líder (al igual que el usuario) desconoce el texto que el sistema le está dictando (así

como el texto que lleva copiado el usuario), limitándose a leer lo que el sistema le va poniendo por pantalla). Cuando el sujeto de prueba interactúa con una configuración humana (C), es el mismo líder quien asume la responsabilidad de dictar el texto, para lo cual el sistema le muestra en su interfaz tanto el texto íntegro que debe dictar, como el texto que el usuario ha copiado. En cualquiera de los tres casos la actitud del líder será la misma, sin que el usuario pueda conocer que existen diferencias entre las labores que desempeña el líder en cada uno de los tres casos. El usuario tampoco conocerá que participa en un proceso de evaluación de la estrategia de toma de turno de un sistema de interacción natural. En su lugar, el experimento se le presentará como un ejercicio para recopilar corpus en el dominio de dictado.

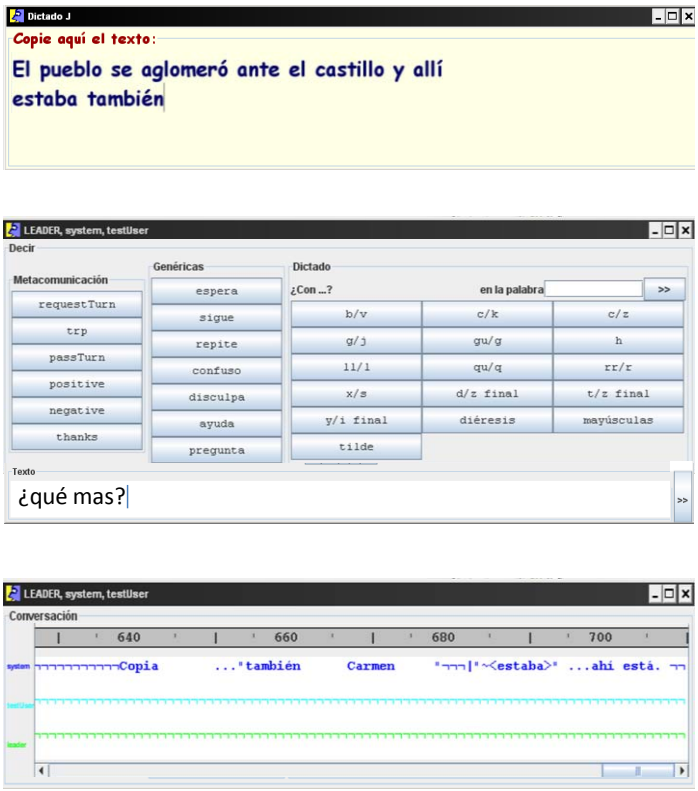


Figura 33: Interfaces gráficas mostrados a cada uno de los participantes en la interacción. De arriba abajo: Interfaz del sujeto de prueba; Interfaz del mecanógrafo; e Interfaz del líder.

7.4 *MÉTRICAS*

Partiendo de los estudios presentados en el estado del arte [Apartado 3.5], la evaluación se compone de parámetros objetivos (para la evaluación técnica) y de parámetros subjetivos (para una evaluación de usabilidad). Los parámetros objetivos contemplan la evaluación de la eficiencia, eficacia y calidad de la interacción, y son extraídos automáticamente durante el

desarrollo de las sesiones. Los parámetros considerados para la evaluación de la eficiencia son: la duración de la sesión (en segundos); el número de contribuciones primarias y secundarias realizadas; y el tiempo de palabra vacante. Para la evaluación de la eficacia se consideran tanto el número de metas desarrolladas como el porcentaje de tiempo de posesión de la palabra de cada uno de los participantes. La medida de la calidad se realiza considerando el número y los tipos de decisiones de toma de turno desarrolladas; el porcentaje de metas resueltas con éxito; y el número de solapamientos e interrupciones ocurridos entre contribuciones durante la interacción. Los resultados de cada una de las configuraciones máquina del sistema (A y B) son promediados y se consideran de forma comparativa entre ellos.

Con el objetivo de evaluar la usabilidad del sistema, se solicita al usuario valorar su satisfacción tras probar cada una de las configuraciones. Esto se realiza sobre una escala tipo Likert [22 ; Figura 34]. Las valoraciones de las distintas configuraciones son mostradas simultáneamente sobre la misma escala para facilitar al usuario ofrecer una valoración global y comparativa de las tres configuraciones. Junto a esto, al final del ejercicio el usuario debe rellenar un formulario [Tabla 11] con el que juzgar la naturalidad de la interacción de cada una de las configuraciones en los siguientes aspectos: dinamismo y organización de la interacción; actitud colaborativa del sistema; adecuado que resulta la estrategia de toma de turno para la resolución de la tarea; cómoda que resulta la interacción con el sistema; y preferencia del usuario sobre alguna de las estrategias frente a otras. Para cada una de las categorías, el usuario marca la mejor configuración con un punto y la peor con cero (asignándose 0,5 puntos a la restante). El formulario también contendrá preguntas sobre el sexo, la edad, la formación y la familiaridad del usuario con los ordenadores, para así poder categorizar a los participantes.

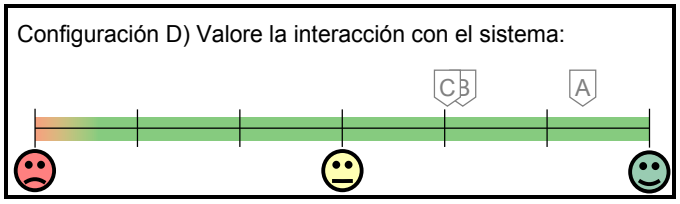


Figura 34: Cuestionario de valoración de la satisfacción del usuario para una configuración de toma de turno dada

Durante los experimentos, también se valora lo humano que resulta para el usuario el sistema con cada una de las configuraciones propuestas. Para ello, el usuario indica en cada uno de los casos si le pareció interactuar con una persona o con una máquina (un punto si le pareció una persona, cero si le pareció una máquina y 0,5 puntos si no fue capaz de identificar si fue una persona o una máquina).

Tabla 11: Cuestionarios de evaluación subjetiva de las distintas configuraciones de toma de turno

Por favor, conteste a las siguientes preguntas:

Pregunta	Respuestas posibles
Edad:	[0..99]
Sexo:	[hombre;mujer]
Formación académica (Indique la que mejor se ajusta a su caso):	[sin graduado escolar;graduado escolar;bachillerato;estudios superiores o universitarios;estudios de postgrado;experto en Tecnologías de la Información;experto en Interacción Hombre-Máquina]
¿Con qué frecuencia utiliza Vd. ordenadores?	[nunca;ocasionalmente;habitualmente]
¿Alguna de las formas de dictar le ha parecido más mecánica que el resto? Si es así, ¿cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
¿Y menos? Si es así, ¿cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
¿Alguna de las formas de dictar le ha parecido más desordenada y caótica que el resto? Si es así, ¿cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
¿Y más ordenada y estructurada? Si es así, ¿cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
¿Bajo alguna de las configuraciones el líder del experimento se ha mostrado más atento y colaborativo que en el resto? Si es así, ¿cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
¿Y menos? Si es así, ¿cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
¿Considera que alguna de las formas de dictar hace más fácil resolver el ejercicio que el resto? Si es así, ¿Cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
¿Y más difícil? Si es así, ¿Cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
¿Durante algún dictado se ha sentido más cómod@ que en el resto? Si es así, ¿en cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
¿Y más incómod@? Si es así, ¿en cuál? (Si duda entre varias, marque "no")	[no;A;B;C]
Si pudiese elegir entre las distintas formas de dictar propuestas, ¿con cuál se quedaría? (Si duda entre varias, marque "ninguna")	[ninguna;A;B;C]
¿Descartaría alguna de ellas? (Si duda entre varias, marque "ninguna")	[ninguna;A;B;C]

En algunos de los casos, quien le ha dictado podría no haber sido una persona, sino una máquina que elaboraba y mostraba por pantalla al líder los mensajes que debía ir leyendo.

Antes de finalizar el ejercicio, indique, para cada uno de los casos, si cree que quien le ha dictado ha sido una persona o una máquina:

Pregunta	Respuestas posibles
Dictado A	[persona; máquina; no sé]
Dictado B	[persona; máquina; no sé]
Dictado C	[persona; máquina; no sé]

7.5 *ANÁLISIS DE LOS RESULTADOS*

Los resultados objetivos se relacionan con una evaluación técnica y funcional de la propuesta. Estos resultados ofrecen información sobre la eficiencia, la eficacia y la calidad de la

propuesta (configuración C) frente a una estrategia de toma de turno desarrollada por ciclo de interacción (configuración B). En lo que respecta a la eficiencia, la Tabla 12 muestra cómo el tiempo necesario para resolver la tarea se reduce de forma muy significativa (en torno a un 42%). Del mismo modo, bajo una toma de turno avanzada es posible resolver la interacción con sólo la mitad de contribuciones (un 48%), proporción que se mantiene tanto para las contribuciones primarias como para las secundarias. Además, la estrategia de toma de turno avanzada hace posible reducir el tiempo que la palabra permanece vacante en más de un 60%, haciendo la toma de turno más ágil y flexible que ante un desarrollo por ciclo de interacción.

En lo que respecta a la eficacia, la propuesta permite doblar el número de metas que son desarrolladas a lo largo de una sesión (aumentando en un 110%). Del mismo modo, el usuario necesita tomar la palabra un 20% menos del tiempo que necesita para alcanzar los mismos objetivos bajo una toma de turno por ciclo de interacción, lo que se hace posible gracias a que el procesamiento incremental permite interpretar sus contribuciones en el momento en que ocurren.

La calidad de la interacción también se ve reforzada bajo una configuración de toma de turno avanzada, puesto que permite revisar con una mayor frecuencia la decisión de toma de turno del sistema (aumentando con la configuración B casi en tres veces el número de decisiones de turno desarrolladas con respecto a la configuración A). Esta mayor frecuencia de revisión de la decisión de toma de turno permite detectar un 15% más situaciones en las que el sistema puede comenzar a contribuir en la interacción y que, bajo un ciclo de interacción, no pueden ser detectadas. Entre estas situaciones se encuentran los casos en los que, estando el usuario participando, el sistema también podría tomar turno sin interferir en la contribución del usuario (por ejemplo, para realizar contribuciones secundarias), o cuando el sistema estaba designado como candidato a tomar la palabra y ocurría una TRP en la contribución del hablante. Bajo una configuración de toma de turno avanzada, el sistema permite identificar de media por sesión 13,5 ocasiones en las que, ante los cambios ocurridos en las circunstancias sociolingüísticas o en el estado de las metas, se requería una revisión de la contribución del sistema (estas situaciones pasaban desapercibidas bajo una toma de turno por ciclo de interacción). Concretamente: la contribución del sistema era adaptada de forma fluida e imperceptible para el usuario en 5,21 ocasiones; se requería su rectificación en 0,34 ocasiones; y la contribución en curso del sistema era interrumpida o cambiaba de rumbo 7,95 veces por sesión. Esta mayor flexibilidad en la generación de las contribuciones reduce el tiempo y el número de contribuciones necesario para resolver la tarea del dictado y, del mismo modo, permite resolver de forma satisfactoria un cinco por ciento más de las metas planteadas a lo largo de la interacción. Finalmente, la toma de turno avanzada también permite un mejor

tratamiento de las situaciones de solapamientos e interrupciones de la interacción. Con ello se mejora la gestión de los solapamientos, reduciéndolos en casi un 43%, y de las interrupciones, que disminuyen en un 44%.

Tabla 12. Resultados de la evaluación técnica para las configuraciones de toma de turno por ciclo de interacción y toma de turno avanzada.

		Ciclo de Interacción	T.T. Avanzada
Duración (segundos)		41,82	17,50
Número de Contribuciones		5,32	2,57
	Primarias (%)	63,91	63,81
	Secundarias (%)	36,09	36,19
Número de Metas		5,68	11,96
	Resueltas (%)	80,49	85,97
	Canceladas (%)	19,51	14,03
Tiempo de Posesión de Palabra (%)			
	Usuario	24,36	19,52
	Sistema	61,23	71,76
	Vacante	14,41	8,72
Número de Decisiones de Toma de Turno del Sistema		18,72	52,32
	No Tomas de Turno (%)	16,83	36,57
	Tomas de Turno (%)	1,89	2,23
	Reformulaciones Suave (%)	0,00	5,21
	Rectificaciones (%)	0,00	0,34
	Auto interrupciones (%)	0,00	7,95
Número de Solapamientos		2,87	1,32
Número de Interrupciones		1,64	0,74

En lo que respecta a la usabilidad [Tabla 13], un análisis inicial de los resultados muestra que la configuración de toma de turno avanzada (configuración B) mejora notablemente la interacción con respecto al ciclo de interacción (configuración A) para los parámetros satisfacción de los usuarios; organización de la interacción; adecuado de la estrategia de toma de turno para la resolución de la tarea; preferencia de los usuarios; y colaboración del sistema.

Tabla 13. Medias (junto con las desviaciones típicas, máximos y mínimos) de todas las configuraciones. *Min = 0 y Max = 1 en todos los casos

	Satisfacción				Organización*		Adecuada*		Cómoda*		Preferencia*		Colaboración*		Dinamismo*	
	M	DT	MIN	MAX	M	DT	M	DT	M	DT	M	DT	M	DT	M	DT
A: Ciclo	2,84	1,92	0	7	0,21	0,30	0,13	0,27	0,13	0,27	0,12	0,24	0,32	0,31	0,44	0,33
B: Advan.	5,45	2,29	0	10	0,46	0,35	0,49	0,31	0,54	0,35	0,53	0,38	0,51	0,29	0,41	0,36
C: Humano	5,94	1,95	3	10	0,71	0,34	0,76	0,30	0,73	0,30	0,69	0,34	0,56	0,31	0,53	0,36

De un Análisis de la Varianza posterior [Tabla 14], los resultados relativos a la organización de la interacción y a lo adecuada y cómoda que resulta la estrategia de toma de turno son significativamente mejores para una toma de turno avanzada que para una el ciclo de

interacción (con p-valores inferiores a 0,005 en todos estos casos). Del mismo modo, la toma de turno avanzada, con respecto a estos mismos parámetros, obtiene resultados significativamente buenos con respecto a la toma de turno humana. La organización media que perciben los usuarios es superior para la configuración por toma de turno avanzada (M = 0,46; DT = 0,35) que para el ciclo de interacción (M = 0,21; DT = 0,30) y es muy próxima a la de la toma de turno humana (M = 0,71; DT = 0,34). La toma de turno avanzada (M = 0,49; DT = 0,31) es, del mismo modo, más adecuada para la tarea que la toma de turno por ciclo de interacción (M = 0,13; DT = 0,27). También obtiene buenos resultados con respecto a la toma de turno humana (M = 0,76; DT = 0,30). Junto a esto, se observa que la toma de turno avanzada (M = 0,54; DT = 0,35) se percibe como más cómoda que el ciclo de interacción (M = 0,13; DT = 0,27) en el dominio propuesto, aunque no tanto como la toma de turno humana (M = 0,73; DT = 0,30).

Tabla 14. Análisis de la Varianza (ANOVA) para todas las configuraciones y pares.

	ANOVA							
	Todas las configuraciones		**Comparaciones por pares**					
	$F_{A,B,C}$	$P_{A,B,C}$	$F_{A,C}$	$P_{A,C}$	$F_{A,B}$	$P_{A,B}$	$F_{B,C}$	$P_{B,C}$
Satisfacción del usuario	25,55	**<0,0001**	50,17	**<0,0001**	29,74	**<0,0001**	1,04	0,311*
Organización de la interacción	22,40	**<0,0001**	48,00	**<0,0001**	12,10	**0,001**	9,71	**0,003**
Adecuado de la estrategia	43,99	**<0,0001**	92,91	**<0,0001**	28,92	**<0,0001**	14,96	**<0,0001**
Comodidad de la estrategia	38,38	**<0,0001**	85,66	**<0,0001**	33,03	**<0,0001**	6,75	**0,011**
Preferencia del usuario	32,63	**<0,0001**	75,44	**<0,0001**	32,37	**<0,0001**	4,21	0,044*
Actitud colaborativa del sistema	6,932	**0,001**	11,99	**0,001**	7,84	**0,006**	0,57	0,452*
Dinamismo de la interacción	1,17	0,315*	1,32	0,255*	0,11	0,743*	1,99	0,162*

La estrategia de toma de turno avanzada mejora significativamente la interacción con respecto al ciclo de interacción para los parámetros satisfacción de los usuarios; preferencia de los usuarios; y actitud colaborativa del sistema. La satisfacción de los usuarios es claramente mayor para las tomas de turno avanzada (M = 5,45; DT = 2,29) y humana (M = 5,94; DT = 1,95) que para el ciclo de interacción (M = 2,84; DT = 1,92). Ambas, toma de turno avanzada (M = 0,53; DT = 0,38) y humana (M = 0,69; DT = 0,34), son también preferidas por los usuarios, en comparación con el ciclo de interacción (M = 0,12; DT = 0,24). La colaboración del sistema mejora en las tomas de turno avanzada (M = 0,51; DT = 0,29) y humana (M = 0,56; DT = 0,31) frente al ciclo de interacción (M = 0,32; DT = 0,31). En todos estos casos, las comparaciones por pares de las tomas de turno avanzada y humana entre ellas no ofrecen resultados significativos. Por otro lado, los análisis relativos al dinamismo de la interacción con respecto a la toma de turno no revelan resultados concluyentes.

En resumen, la toma de turno avanzada se comporta de una forma notablemente más natural que la toma de turno por ciclo de interacción (aunque en el caso del dinamismo no se obtuvieron resultados significativos). En la misma línea, las diferencias encontradas son muy claras al comparar el ciclo de interacción con el resto de configuraciones de toma de turno, pero no lo son tanto cuando se comparan entre ellas la toma de turno avanzada y la toma de turno humana.

En lo que respecta al parecido humano del sistema con cada una de las tomas de turno [Tabla 15], se observa que la mayoría de los usuarios identificaban la toma de turno por ciclo de interacción con un comportamiento máquina (21 usuarios, frente a 16 que creían que se trataba de una persona y dos que no sabían). Junto a esto, los usuarios tendían a considerar que se encontraban hablando con otra persona cuando se trataba de la toma de turno avanzada (29 usuarios, comparados con seis que creían que se trataba de una máquina y cuatro que no sabían) y la toma de turno humana (28 usuarios, frente a siete que creían que se trataba de una máquina y cuatro que no sabían). Un Test de Independencia por $\chi 2$ permitió determinar que las diferencias encontradas son claramente significativas en un análisis general de las tres configuraciones. En la comparación por pares, los resultados obtenidos para el ciclo de interacción son significativamente peores que para el resto de configuraciones. Los resultados en la comparación entre la toma de turno avanzada y la toma de turno humana no ofrecen resultados concluyentes. En consecuencia, se puede afirmar que los usuarios detectan claras diferencias entre una toma de turno por ciclo de interacción y el resto de las estrategias de toma de turno, pero que no las detectan o estas no están tan claras cuando se comparan la toma de turno avanzada y la toma de turno humana entre ellas.

Tabla 15. Izq.: Usuarios que identifican el sistema como máquina o persona para cada una de las configuraciones. Derch.: Test de Independencia $\chi 2$ para todas las configuraciones y pares.

	Parecido humano			**Test de Independencia $\chi 2$**							
				Comparación general		**Comparaciones por pares**					
	Máqui-na	**NS / NC**	**Huma-no**	$\chi^2_{A,B,C}$	$P_{A,B,C}$	$\chi^2_{A,C}$	$P_{A,C}$	$\chi^2_{A,B}$	$P_{A,B}$	$\chi^2_{B,C}$	$P_{B,C}$
A: Ciclo	21	2	16	17.51	**0.002**	10.94	**0.004**	12.76	**0.002**	0.09	0.954*
B: Advanzada	6	4	29								
C: Humana	7	4	28								

Capítulo 8 CONCLUSIONS AND FUTURE WORKS

Once all the stages of this thesis have been accomplished, some of the conclusions obtained are presented, and we describe some of the future lines that follow to this work.

8.1 *CONCLUSIONS*

This study introduces a novel architecture for Natural Interaction Systems that allows advanced management of turn-taking in human-computer interactions. This architecture is adapted to carry out more natural turn-taking to the one that people perform in their human interactions. This architecture includes the new components Continuity Manager, Processes Coordinator and Turn-Taking Manager to address the key issues involved in advanced turn-taking. Similarly, a Dialogue Manager apt for application under this advanced turn-taking is described. This Dialogue Manager makes the reinterpretation of the input contributions possible, and enables the system for performing independent processes of formulation of contributions and updating of the state of the interaction after they have been expressed. The system represents the primary or collateral nature of contributions, and it estimates the commitment established regarding the development of the goals of the interaction at any time - and the cost-benefit relationship of carrying them out -.

The turn taking in human interaction has been a subject widely discussed from the disciplines of linguistics, sociology and sociolinguistics. In particular, ethnomethodology, and more specifically conversation analysis, has paid particular attention to the dynamic way in that

language is used by people in their spontaneous interaction. These studies consider the interaction that people realize as a joint action. According to them, the actions that participants perform within the interaction are agreed and committed by all participants through their individual actions. In other words, participants take part in the interaction with the aim of achieving their own goals and they assume as a counterpart the need of facilitating the development of their interlocutors' goals through the interaction. As a result, participants commit to a greater or lesser extent their interlocutors' goals, fact which creates an abstraction - usually called common ground - which is composed of the joint goals that participants create around their individual goals. All actions in the interaction are, in themselves, joint actions. This includes any contributions made by the participants during the interaction which, according to the principle of joint construal, are built jointly by both the speaker and his interlocutors through the simultaneous feedback that they offer constantly. Similarly, the distribution of turns is also a joint action that arises from the confluence of the participants' interests, which take place in particular sociolinguistic circumstances that evolve over time.

In this turn-taking there is no a predefined order of intervention for the participants, and neither the duration nor the content of their turns are defined unilaterally by the participant who expresses them. In human interaction, the speaker is not the participant who gives the floor to the next speaker when he considers that his intervention has ended. In fact, the possession of the floor is reviewed in a joint way by all participants in those parts of the intervention that can be considered Transition Relevant Places (TRPs). This revision of the possession of the floor is based on the participants' conjectures about the state of the turn of each participant, who is the current speaker, and the participants that are posing as candidates to take the floor (by explicit or implicit appointments, expressed by the current speaker or by the candidate own request). In any case, not all the participants' contributions are subject to these turn-taking rules. Both the secondary track contributions (which aim to develop metacommunicative aspects of the interaction) and the contributions which do not distract the participant's attention away from the speaker's primary contribution (due to they can be expressed through alternative modalities) can be expressed in the interaction without require to be in the possession of the floor. Similarly, the particular interests of the participants, the sociolinguistic circumstances that surround the interaction and even errors that occur on the conjectures on which participants base their turn-taking decisions can trigger struggles for the possession of the floor and situations of overlap or interruptions. These situations, far from being anomalous, are necessary and frequent resources in the human turn-taking. In short, in the human interaction participants are subjected to the constantly review of their position on the turn taking (not just after the end of each contribution, but also during its development) and they must estimate the turn-taking state, the state of the

goals of the interaction and the sociolinguistic circumstances under which the interaction is carried out, in order to perform the turn-taking decisions. As a result of this decision can be both primary (interventions) as secondary contributions (e.g. simultaneous feedback), and these will be produced by any of the available modalities. Similarly, participants are constantly forced to reformalize and adapt their ongoing contribution depending on the evolution of the turn-taking state, the interactive goals and the sociolinguistic circumstances of the interaction. Participants are not limited to express their contributions as they were initially formalized.

Natural Interaction Systems are a set of systems that Human-Computer Interaction has addressed in order to make the technology accessible to people through the same codes, modalities and procedures that people use in their human interaction. Such systems have made substantial progress in recent years and they are currently applied on a wide variety of domains, among which are: customer service, education, assistance to users, or video games. There is great diversity of natural interaction systems, which can be classified based on different criteria: i) depending on the role played by the system as an interactive agent, systems can be task-oriented, conversation simulation or discursive; ii) depending on the modalities and codes used, they can be monolingual, multilingual or multimodal; iii) by the direction of the initiative, they can be user- or system-guided or mixed initiative; and iv) depending on how they model the dialogue, there are dialogue grammars, frames, intentional models, joint-action models and statistical models. Generally, interaction systems tend to simplify turn-taking considering it a pass-the-baton process by which the floor is passed from one participant to another in an organized manner, and in which only the person in possession of the floor can contribute to the interaction and determines unilaterally how he will do so. In this way, interaction develops from beginning to the end as a cyclical process, entitled the interaction cycle.

In order to develop advanced turn-taking, discursive systems (such as TRIPS or Jaspis systems) are relevant to consider because the manner and moment in which they produce turns during the interaction are derived from their in-depth comprehension, as well as from the surrounding circumstances. Multimodal systems (MATCH and SmartKom, among others) are also relevant to consider because natural-interaction turn-taking is largely managed by paralanguage. A mixed initiative system (e.g. TRIPS or CMU) is of interest because in natural interaction, the initiative to add, eliminate or carry out goals can arise at any moment and come from any of the participants. Finally, solutions based on the theories of joint action (for example, TRIPS or SOPAT) are very close to human interaction because it is the balance between the goals of the different participants and the compromise reached that determines with what urgency the participants should take part in the interaction and what goals should be developed when doing so.

Among the proposals that deal with the resolution of some of the turn-taking issues inherent in dialogue, barge-in systems stand out. These systems are capable of managing the interruption of their own contributions when the user starts a new utterance simultaneously. Systems with backchannel management also stand out. These systems understand the need to offer the user secondary track uptakes in order to create a joint construal of what he is taken to mean. Even if in these cases the moments in which the system should offer backchanneling are determined by factors such as pitch, prosody, gaze movements or silences, and none of these systems considers for this purpose the level of commitment reached concerning the goals, the state of turn-taking or the influence of socio-linguistic circumstances. The conclusions are not applicable to the estimation of the moments in which the system can participate by making any type of contribution, be it primary or secondary, nor do they consider the effect that the user backchannel (among other types of contribution) has on the system's own contributions.

Systems that require incorporate the effects of the user backchannel in the ongoing system's own contribution must deal with the problem of the incremental processing. Incremental processing is the successive processing of interpretation and generation in increments smaller than whole utterances, as it can be observed in natural interaction. Other systems, such as VM-GEN and Max, consider the user's simultaneous backchanneling by means of an incremental approach. There are also other authors that have dealt with the problem of incremental processing. The Ymir system detects turn-taking markers present in the user's contribution (silences, allowing the other to speak etc.) and applies them to a turn-taking decision process, although limiting itself to estimating when the user is waiting for the system to take its turn, without developing at any time active turn-taking based on the theories of joint action. The FADE system complements a turn-taking decision, which is also passive, with incremental processing. None of these systems tackles identifying those situations in which the system should take, keep or give the turn on its own initiative beyond the expectations of the user. In this sense, the DECOP system evaluates the cost benefit relationship of intervening, and applies it to a turn-taking decision from which it determines if it should interrupt the user or not (although without making a distinction between primary track interventions and secondary track uptakes and without applying that decision to incremental processing).

In short, no system tackles in an integrated way natural turn-taking that combines incremental processing (which makes co-operative construction of all the users' and system's contributions possible) and performs joint-action turn-taking decision in which both the primary or secondary nature of the participant's contributions - and the benefit-cost relationship of participating in the action - are taken into consideration. Such a turn-taking process would allow the system to take an active part in the sharing-out of turns (in the same way that human

interlocutors do) and would allow the user to know that the system assumes this responsibility and then interact with it, as he would do with a human interlocutor.

This thesis proposes architecture of Natural Interaction System with advanced turn-taking abilities. This proposal aims to fill some gaps found in the state of the art in regard to the temporal development of the interaction. This architecture consists of a multimodal discursive joint-action system with mixed initiative. It is implemented on a blackboard-oriented, multi-agent platform. Its knowledge models are implemented in one or more agents running on a standalone machine or distributed through a LAN. Agents offer services to one another, and results usually come from the collaboration of the whole. The proposal takes as its starting point previous works done in the Advanced Databases Group of the University Carlos III of Madrid in the area of Natural Interaction, that describe an architecture composed of Ontology, Interface Components, Interaction Manager and Presentation Manager. This dissertation includes new skills and components to this architecture in order to address the specific problems of advanced turn-taking.

Among these new skills are: development of reinterpretation processes; separation of the formulation of contributions and updating the state of the interaction after they have been expressed; representing the primary or collateral nature of contributions; and the estimation of the commitment established regarding the development of the goals of the interaction and the cost-benefit relationship of carrying them out at any time. These new features are provided by the Interaction Manager, and more specifically, by the Dialogue Manager component. In addition to them, Continuity Manager, Coordination Manager and Turn-taking have been included in the Presentation Manager to the architecture.

The Continuity Manager is the component in charge of the incremental treatment of the processes of interpretation and generation. It involves enabling the system to carry out interpretations and output contributions in real time, but also identifying and filtering disfluencies in the user's contributions (hesitations, discontinuities, omissions, repetitions or rectifications) and generating them in the system's. In order to do so it maintains a partial representation of user contributions which is updated every time it receives new fragments of user contributions from the Interface Components (fragments where granularity can be treated at token or n-grams levels, but also at regular periods of time). When this occurs, the Continuity Manager combines the new fragments of contribution with the user's partial contribution in larger units to detect portions of contributions that, by themselves, represent complete communicative acts. When new complete communicative acts in the user's contribution are identified, they are sent to the Dialogue Manager in order to interpret which developments they

mean in the interaction state and in the associated sociolinguistic knowledge. Regarding the generation of the system's contributions, the changes that occur in the socio-linguistic circumstances surrounding the interaction, or those in the state of the interaction itself, motivate the Dialogue Manager to review the state of turns. This occurs when there are generation requests made by the Turn-taking Manager after its turn-taking decisions processes, which can trigger formulation, reformulation, rectification or self-interruptions of the system's contribution. The Continuity Manager is in charge of combining these new fragments of contributions with the ongoing flow of system expression. Hesitations, discontinuities and other disfluencies are common phenomena in spontaneous speech. Some of them take place as consequence of improvisation, adaptation of the speaker to the changing circumstances or errors. Some others are used by the speaker as linguistic resources (working as turn-taking markers).

For its part, turn-taking management falls on the Turn-taking Manager component. This component represents the activity state of the turn of each participant (if he is currently expressing a contribution or he is not); the primary or secondary nature of these turns; and stores the estimation of the Transition Relevant Places, or TRPs, achieved in them. Similarly, this component maintains a conjecture concerning which participant is considered to be the current speaker and which are presented as candidates to take the floor (either because they requested it or by designation by the current speaker). Updating the turn-taking status is invoked by the Continuity Management and Dialogue Management components when turn-taking acts are identified in a participant's contribution, or when the state of the interaction or the socio-linguistic circumstances change during the processes of interpretation and generation. Starting with this knowledge, and considering the status of the goals of the interaction and the associated socio-linguistic knowledge, the Turn-taking Manager develops the system's turn-taking decision processes. With them, the system decides whether or not to contribute to the interaction, and determines what goal or goals it should develop to do so. In order to perform a turn-taking decision, the Turn-taking Manager considers the following parameters: the urgency of making progress for each goal; if these goals are primary or secondary (for example, backchannel); if the system has the floor or not; if it is or is not a candidate to take the floor; and the state of the participants' turns (if they are carrying out any activity, if they are silent, etc.). These processes are triggered by the changes produced in the participants' state of turns, possession of the floor, candidates to take the floor and the goals of the interaction. Therefore, turn-taking decisions are processes requested by the Turn-taking Manager and the Dialogue Manager. The turn-taking decision starts with the identification of the goals that the system can make progress in the interaction (at this moment) and it ends with a new generation request to

the Dialogue Manager (in which the final set of goals that will be developed in the interaction at this precise instant is identified).

Finally, the simultaneous development of the interpretation and generation processes makes it necessary to manage access to the shared resources of the system (knowledge about the state of the interaction, the session, the situation, users, emotions, and self model) and to systemize the order in which they are executed in order to guarantee that the maximum tolerated waiting restrictions - with which the interlocutors expect to receive backchannel on their actions - are fulfilled. The component in charge of this function is the Coordination Manager.

With all this, the proposed Natural Interaction System allows the development of a wide range of possibilities regarding the formalization of the system's contributions:

- Initial formalization: The system was not previously developing a turn and a contribution has been produced as a result of the formalization.
- Reformulation: The system was previously developing a turn and the expression of the new contribution can replace the previous expression without involving a break in the flow of the system's contribution.
- Rectification: The system was developing a turn and the new contribution corrects part of what the system was expressing in its current contribution.
- Self-interruption: The system was developing a turn, and the new contribution involves a break with the contribution that they had developed so far (or forces its end).

With respect to the turn-taking decision, it has been raised from a joint action point. This enables the system to take an active role in the distribution of turns, in the same way that human participants do in the interaction. This approach allows the system to address a wide range of turn-taking situations:

- To take turn to carry out a secondary contribution, regardless of the turns state, the possession of the floor, the candidates to take it or the urgency.
- To take the turn if the system has something to say and the floor is vacant.
- To take the floor if the system has something to say and no participant is developing a primary turn.
- To try to take the turn if the system has something to say, there is a speaker, the speaker has reached a TRP in his contribution and the system is candidate to take the floor because of an speaker appointment (with a higher priority if it has been designated and lower if he request it) .

- To request the floor if the system has something to say, the floor is not vacant, the speaker's contribution has not reached a TRP and the system deems urgent to speak.
- To interrupt, even if there is another speaker and the system has not reached a possible TRP when the system deems urgent to speak and the sociolinguistic circumstances allow the system to speak (for example, as a matter of dominance of their role compared to the role of the current speaker).
- To interrupt regardless of whether there is or not speaker and of the sociolinguistic circumstances if the system deems of a maximum urgency to speak (for example to warn of hazards).

Similarly, if the issue is to decide whether the system should continue performing its current turn, the following scenarios are included:

- To continue performing the contribution if the system is the speaker and there is no participant performing contributions that could entail interference in its contribution (e.g. expressing just secondary contributions).
- To continue performing the contribution if it is a secondary contribution, even when other participants may be performing contributions at the same time.
- To finish its contribution if the system is the speaker, it has reached a TRP and other participants try to take the floor (as long as the sociolinguistic circumstances or urgency of the goals do not dictate otherwise).
- To reformalize its current contribution when they occur changes in the state of interaction, in the interactive goals or in the sociolinguistic circumstances that directly affect the contribution that the system is carrying out (that may result in soft-reformulations, rectifications or self-interruptions).

This proposal has been implemented and evaluated in the LaBDA-Interactor framework. This evaluation is based on real interactions maintained between the system and test users. These interactions were restricted to a specific interaction domain and they took place under different configurations of the system. For the evaluation of the proposal, different interaction domains were analyzed. In them, the skills required for advanced turn-taking were put into play and they were chosen with the aim to minimize the influence of the problems associated with the non analyzed levels of interaction (such as voice recognition, speech synthesis, natural language processing and multimodal adaptation). The selected interaction domain was the “dictation domain”. In this domain the system played the role of a boss (the participant who dictates a text), and the users the role of secretary (the participant who copy the

text dictated by the boss). The dictation domain was very suitable for evaluating advanced turn-taking because it includes primary and secondary contributions and long contributions subjected to overlaps, interruptions and changes of speaker. This domain includes dialogs where the system corrects spelling or typographical errors in the text copied by the user, the user consults the system about spelling or user request help, repetitions or waits, among other possible dialogs.

Both participants, user and system, share the major goal of completing the dictation, taking care of aspects such as orthography, punctuations marks and the case of characters. Each test subject completed three different dialogues with the system, each one with a different interaction style. More specifically, user and system complete three different dictations with three different turn-taking configurations. These turn-taking configurations were: an interaction cycle (turns are handled sequentially in a predefined order of participation); advanced turn-taking (the system puts into practice the strategies proposed in this work); and Wizard of Oz (a human determines what the system has to say and how it has to behave at all times). These configurations were presented to the user in random order (different for each participant), so that they were not aware of which configuration was being applied. Furthermore, they were also unaware of which configuration was under evaluation (if any), and the parameters considered in the evaluation.

The measurement of results was based on both objective (technical evaluation) and subjective parameters (usability evaluation). The set of technical parameters includes the percentage of goals solved and cancelled, the percentage of time of possession of the floor for each participant, the percentage of time the floor was vacant or the number of turn-taking decisions performed. In order to evaluate usability, users were required to value their satisfaction with each configuration and, after all three dictations, they filled out a questionnaire form that contained questions to ascertain which configuration produced a more (and less) dynamic, organized, collaborative, suitable and comfortable interaction, and if the user preferred (or rejected) any of them.

The results of this evaluation reveal that users value the presented turn-taking strategy as clearly more natural and human than the classic interaction cycle approach, although there is still room for improvement with respect to real human behaviour. It obtains marks very close to human interactive abilities in aspects such as organization, collaborativeness, suitability and comfortability for certain interaction domains (where turn-taking plays an important role).

8.2 *FUTURE WORKS*

Although this work provides the basis for the treatment of the turn-taking markers produced as alterations in the temporal continuity of the ongoing contribution through its Continuity Manager component (both on the interpretation and in the generation sides), and although some of these markers have been dealt in this work (mainly silences and filled silences), most of them have not been covered during the course of this thesis. Its application for gaining, maintaining and transferring the attention could be of a great interest. Some types of disfluency in the participants' contribution (such as repetitions, hesitations or corrections) have also been treated. Nevertheless, most of them have not been addressed. The treatment of other markers of turn taking, such as disfluencies, could be tackled considering the continuity management skills introduced in this work. Given its importance to the natural development of turn-taking, it is contemplated to continue the investigation in this line as a future work. With respect to the turn-taking markers that are explicitly expressed by the participants verbally or through paralanguage (gestures, looks, etc.), they have only been treated at the level of communicative act, and its acquisition and synthesis have been simulated by buttons and graphic components in a user interface. Thus, the interpretation and natural language generation of these markers has not been treated by now.

On the other hand, a Turn-Taking Manager has been presented. This component makes it possible to estimate both the state of possession of the floor, and the participants that are candidates to take it. This component applies these conjectures to perform the turn-taking decisions and these decisions are triggered by changes in the status of the turn-taking and in the goals of the interaction. Thus, the system can decide when to participate in the interaction, when it should continue performing its contribution or when to stop. All of this in a flexible way based on the rules described by the turn-taking system of Sacks et al. Moreover, the model includes the modulation of these decisions based on the interaction state, the individual and joint goals, and the circumstances that surround the interaction, so the model is presented as domain independent. In any case, it is intended to deal with the separation of the model and the set of rules applied to the turn-taking decision in order to make the proposal configurable and to adapt it to other turn-taking systems different from that Sacks et al. propose.

The combination of the improvements proposed in this paper, in regard to incremental processing and the turn-taking decision, in combination with some other models of knowledge which are not dealt in this work (as Situation Model, User, Self-model or Ontology) could enhance the interaction and impact in a very favorable way the turn-taking decision. The Situation Model would apply, for example, political aspects (such as the dominance of certain

roles over others) on the turn-taking decision. This component would also include the spatial and temporal aspect of interaction, considering even a management of events that would impact directly on the criticality of the goals and that would improve the management of situations of emergency (giving, for example, more priority to notices of actual risks for the user, such as fires or disasters, than to other issues of interaction). Similarly, the User Model enables the system to adapt the turn-taking decision to the users' preferences, or give priority to some users with respect to others. It would also allow identifying the participant's specific needs and improves the way they are developed through proper goals of recommendation and suggestions -whose development would be managed through an advanced turn-taking strategy-. The Self-model, component that represents the system's emotional long-term-goals and its operational goals, would also benefit from better turn-taking that handles the moments and the way in which they should be performed in the interaction.

Multimodal adaptation, both as to fission and fusion refers, has barely treated along the present work. The proposed architecture provides the location that this component would have in a complete Natural Interaction System, and describes which services it should offer to the rest of components and how it should be communicated with them. Human interaction is developed in a coordinated manner through the different modalities available in the interaction, which may differ depending on the users and the devices used to access to the system. Much of the collateral communication (which is essential on both the joint construal reached by participants, and the turn-taking distribution in the interaction) is performed in a coordinated manner across different modalities, as demonstrated by numerous studies in the area of prosody, intonation, pitch and gestural expression. Without an adequate adaptive multimodal processing this collateral communication cannot carry out in the interaction, with the consequent loss of effectiveness of the turn-taking strategy developed by the system. The integration of the proposed models with a multimodal adaptive component would allow the system to take full advantage of the turn-taking system's skills. This is the objective of another future line of research.

Although the components in which the turn-taking competences fall are already addressed in this work (Continuity Manager, Process Coordinator, Turn-taking Manager and Dialogue Manager), the development of an advanced turn-taking also affects the Interface Components on which an advanced turn-taking development imposes several important restrictions. The Physical Interface Components, both input and output, requires a temporal processing granularity lower than the maximum response delay that people tolerate in the interaction without losing the perception that the interaction takes place in real time (fixed by some studies around 150 and 400 ms.). This restriction also includes the delay with which the

Interface Components realize the acquisition and synthesis of the natural expressions. Similarly, the Physical Output Interface Components must allow confirming the expressed fragments of the system's contribution and they must be capable of canceling the ongoing expressions when needed. The Continuity Manager demands these confirmations in order to determine when each individual communicative act can be supposed as expressed, a prerequisite for the correct incremental processing of the interaction. Regarding the cancellation of the current expression, it is required from the Interface Components the ability to interrupt the system's current expression, contemplating the ability to confirm in advance those fragments of expression whose synthesis is expected during the time interval that the system would required to handle a hypothetical reformulation of the current contribution, in order to ensure the greatest fluidity possible in the system's contribution (and minimize its discontinuities).

With respect to the Logical Interface Components, its operation will be coordinated by the Continuity Manager (component that integrates fragments of contribution in system's complete participations), The turn-taking decisions will be favored where the extent to which these components are able of projecting in advance the end of the communicative acts in the input contributions and of estimating the points of the output expressions in which the interlocutors could project the end of the system's communicative acts. This early management of projections would mean to obtain early predictions of the TRPs (as human participants of the interaction also do) and reach a better compromise with them around the turn-taking state. The early projection of communicative acts improves the utilization of time in the interaction, reducing the existing silence during the floor transfers and improving the choices of the candidates to take the floor.

It is also considered as a future line of work the application of the implemented system to other domains of interaction that can benefit from advanced-turn taking strategies. These domains could be guidance services, entertainment, education and assistance to users, although in general terms, any domain that requires a high degree of pro-activity of the system (or contemplates changes in the sociolinguistic circumstances) would be highly benefited from the system's new skills.

Capítulo 9 CONCLUSIONES Y LÍNEAS FUTURAS

Una vez abordadas todas las etapas del desarrollo de este trabajo de tesis doctoral se presentan algunas de las conclusiones obtenidas y las líneas futuras de investigación que resultan del mismo.

9.1 *CONCLUSIONES*

Con esta tesis doctoral se ha propuesto, implementado y evaluado una arquitectura de Sistema de Interacción Natural capaz de desarrollar una toma de turno más cercana a la que desarrollan las personas durante su interacción humana. La arquitectura incluye los nuevos componentes Gestor de Continuidad, Coordinador de Procesos y Gestor de Toma de Turno para abordar los principales problemas implicados en una toma de turno avanzada a nivel de presentación. Del mismo modo, se revisa el componente Gestor de Diálogo para adaptar su estado de interacción a un modelo con control de versiones e incorporar en él las nuevas habilidades de gestión de metas y gestión de la pista de acción que permiten dar soporte a estos nuevos componentes.

La toma de turno en la interacción humana ha sido una materia ampliamente analizada desde las disciplinas de la lingüística, sociología y sociolingüística. En concreto, la etnometodología, desde su análisis conversacional, ha prestado especial atención a la forma en la que el lenguaje es usado por las personas de forma dinámica en su interacción espontánea. Para ello, parte de la suposición de que la interacción es, en sí misma, una acción combinada.

Los participantes toman parte en ella con el objetivo de satisfacer sus propias metas individuales, pero asumen que para conseguirlo deben facilitar al resto de participantes el desarrollo de sus propias metas, por lo que se comprometen en mayor o menor medida con su desarrollo y, con ello, se establece una zona de común de metas combinadas en torno a sus metas individuales. Todas las acciones de la interacción son, en sí mismas, acciones combinadas. Esto incluye cualquiera de las contribuciones que realizan los participantes durante la interacción y que, según el principio de la interpretación combinada, son construidas conjuntamente tanto por el hablante como por sus oyentes a través de la realimentación simultánea que le ofrecen constantemente. La propia toma de turno que se desarrolla en la interacción humana surge de la confluencia de los intereses particulares de los distintos participantes en unas circunstancias sociolingüísticas concretas que evolucionan con el tiempo.

En esta toma de turno no existe un orden predefinido de intervención de los participantes y, ni la duración ni el contenido de sus turnos, quedan definidos unilateralmente por el participante que los expresa. En realidad, no es el hablante quien cede la palabra a un hablante siguiente cuando considera que su intervención ha finalizado, sino que la posesión de la palabra es revisada de forma combinada por todos los participantes en aquellos puntos de la intervención del hablante que conjuntamente son considerados lugares de transición pertinentes de la palabra (los denominados TRPs, del inglés Transition Relevant Places). Estas revisiones de la posesión de la palabra se apoyan en las conjeturas que realizan los participantes sobre el estado de la toma de turno, que incluye el conocimiento sobre quién es el considerado como hablante actual, quienes son los participantes candidatos a tomar la palabra (bien sea por designaciones explicitas o implícitas del hablante actual, o por la propia solicitud del hablante), cuáles de ellos están desarrollando turno en ese momento, y cuál es el estado en el que se encuentran dichos turnos. En cualquier caso, no todas las contribuciones de los participantes están sujetas a estas reglas de toma de turno y, tanto la contribuciones de carácter secundario (que pretenden desarrollar aspectos metacomunicativos de la interacción), como las que no suponen una desviación de la atención sobre la contribución primaria que desarrolla el hablante (por poder ser expresadas a través de modalidades alternativas), pueden ser desarrolladas en la interacción sin ser requerida para ello la posesión de la palabra. Del mismo modo, los propios intereses particulares de los participantes, las circunstancias sociolingüísticas, e incluso los errores producidos sobre las conjeturas en las que los participantes basan su decisión de toma de turno, pueden desencadenar situaciones de lucha por la posesión de la palabra, solapamientos o interrupciones. Estos, lejos de ser fenómenos anómalos, constituyen recursos necesarios y frecuentes en la toma de turno humana. En definitiva, durante la interacción humana los participantes se ven obligados a revisar constantemente su postura frente a la toma de turno (no

sólo tras la finalización de cada contribución, sino también durante su desarrollo) y, para tomar esta decisión, deben conjeturar el estado de la toma de turno, el de las metas de la interacción y también el de las circunstancias sociolingüísticas bajo las que se desarrolla. De esta decisión pueden resultar tanto contribuciones primarias (intervenciones) como secundarias (por ejemplo, las contribuciones de realimentación simultánea), y estas serán producidas a través de cualquiera de las modalidades disponibles. Del mismo modo, los participantes se ven constantemente obligados a reformular y adaptar su contribución en curso en función de cómo evolucionen este estado de la toma de turno, las metas interactivas y las circunstancias sociolingüísticas de la interacción. En la interacción humana, los participantes no se limitan a expresar su contribución tal y como fue formalizada inicialmente.

Los Sistemas de Interacción Natural son el conjunto de sistemas tratados por la Interacción Hombre Máquina que tratan de hacer accesible a las personas la tecnología a través de los mismos códigos, modalidades y procedimientos que las personas utilizan en su interacción humana. Este tipo de sistemas han experimentado un espectacular desarrollo en los últimos años, siendo en la actualidad aplicados a gran diversidad de dominios, entre los que se encuentran: la atención al cliente; la educación; la asistencia a usuarios; o los videojuegos. Existe gran diversidad de sistemas de interacción, que según las clasificaciones más extendidas pueden ser: orientados a tarea, conversacionales o discursivos (en función del rol desempeñado por el sistema en la interacción); monolingües, multilingües o multimodales (en función de las modalidades y códigos soportados); de interacción dirigida por el usuario, por el sistema o mixta (en función de qué participante toma la iniciativa en la interacción); y basados en gramáticas de diálogo, marcos, sistemas intencionales, sistemas de acción combinada o estadísticos (en función de la estrategia de modelado de diálogo que implementan). Por lo general, todos los sistemas de interacción desarrollan una toma de turno basada en el denominado ciclo de interacción. Según ésta, la palabra es pasada por orden de un participante a otro durante toda la interacción y, en ella, el participante en posesión de la palabra es quien decide cuáles serán los contenidos desarrollados en su contribución y durante cuánto tiempo la tomará.

Para hacer posible un desarrollo de toma de turno más natural, similar al desarrollado en la interacción humana, se requieren sistemas discursivos, multimodales, de iniciativa mixta y de acción combinada. Discursivos (como los sistemas TRIPS o JASPIS) porque la decisión de la toma de turno surge una profunda comprensión de la interacción y de las circunstancias que la rodean. Multimodales (como MATCH o SmartKom) porque la toma de turno de la interacción humana es gestionada en gran medida a través del paralenguaje. De iniciativa mixta (como TRIPS o CMU Communicator) puesto que cualquiera de los participantes debe estar en

igualdad de condiciones para tomar la iniciativa de incluir, eliminar o hacer progresar las metas de la interacción. Finalmente, de acción combinada (como TRIPS o SOPAT), puesto que es el equilibrio entre las metas de los diferentes participantes, y los compromisos que se alcanzan en torno a ellas, los que determinan en última instancia la urgencia con la que los participantes necesitan tomar parte en la interacción (y qué metas desarrollarán al hacerlo).

De todos los sistemas de interacción natural que existen, algunos de los que abordan aspectos de una toma de turno avanzada son los sistemas con gestión de la interrupción, que son capaces de interrumpir su propia contribución cuando detectan que el usuario ha comenzado a hablar. No obstante, esta interrupción se produce sin una gestión de los fragmentos de contribución ya expresados y limitándose a ceder la palabra de forma pasiva, sin evaluar si es lo adecuado según el estado de la toma de turno. También destacan los sistemas que implementan una gestión de la realimentación. Estos entienden la necesidad de ofrecer una comunicación colateral simultánea a la intervención del usuario con el objetivo de simular el establecimiento de una interpretación combinada. Consideran que los únicos factores determinantes en la generación de la realimentación son aspectos como la prosodia, los silencios o los gestos, pero no prestan atención ni al estado de las metas, ni al compromiso alcanzado sobre ellas, y tampoco al estado de la toma de turno o al de las circunstancias sociolingüísticas que envuelven la interacción. Otros sistemas relevantes son VM-GEN y Max, que consideran la realimentación simultánea de usuario por medio de una interpretación incremental. También Ymir, que detecta marcadores involucrados en la decisión de la tomar de turno y los aplica a un proceso pasivo de decisión de toma de turno en el que el sistema estima cuándo el usuario espera de él que tome la palabra. Junto a ellos, FADE complementa una decisión de toma de turno similar a la desarrollada por el sistema Ymir, con un procesamiento incremental. En cualquier caso, ninguno de ellos contempla la identificación de las situaciones en las que sistema debe tomar, mantener o ceder turno por su propia iniciativa, más allá de las expectativas del usuario. A este respecto, DECOP evalúa la relación entre el beneficio y el coste de contribuir en la interacción y desarrolla una decisión de toma de turno que permite determinar si interrumpir o no al usuario, aunque sin hacer distinciones entre contribuciones primarias o secundarias y sin aplicar en su decisión un procesamiento incremental. En definitiva, no existen sistemas que aborden de forma íntegra todos los problemas involucrados en una toma de turno natural, en la que se considere tanto un desarrollo incremental de los procesos de interpretación y generación, como una decisión de toma de turno basada en las teorías de la acción combinada, que contemple de igual modo tanto contribuciones primarias como secundarias, y que permita que el sistema tome parte activa en el reparto de los turnos, al igual que hacen los participantes humanos en la interacción.

En este trabajo se ha propuesto una arquitectura de un Sistema de Interacción Natural con capacidades avanzadas de toma de turno. Esta propuesta tiene como objetivo cubrir los vacíos encontrados hasta el momento en el estado del arte en lo que respecta al desarrollo temporal de la interacción. La arquitectura propuesta puede clasificarse como un sistema discursivo, multimodal, de iniciativa mixta y de acción combinada, que parte de los trabajos desarrollados en el Grupo de Bases de Datos Avanzada de la Universidad Carlos III de Madrid en el área de la Interacción Natural. El sistema está estructurado en un conjunto de componentes autónomos y de procesamiento independiente ideados para funcionar sobre una plataforma multiagente de pizarra compartida. La propuesta toma como punto de partida una arquitectura compuesta de Ontología, Componentes de Interfaz, Componente de Interacción y el Gestor de Presentación, y la completa con nuevas habilidades y componentes para abordar los problemas específicos de una toma de turno avanzada. Entre ellos se encuentran las habilidades de un control de versiones sobre el estado de interacción, la gestión de la pista de acción de las metas desarrolladas y un subcomponente Gestor de Metas, todos ellos como parte del Gestor de Diálogo del Componente de Interacción. Del mismo modo, han sido incorporados al Gestor de Presentación los nuevos componentes Gestor de Continuidad, Gestor de Toma de Turno y Coordinador de Procesos.

El control de versiones del estado de interacción ha sido propuesto con el objetivo de separar los subprocesos de formalización de contribuciones de sistema y la confirmación de los progresos que la expresión de las acciones comunicativas que las componen supone en el estado de interacción. Esta separación se ha llevado a cabo con el fin de hacer posible la reformulación de las contribuciones del sistema. El control de versiones permite recuperar versiones anteriores del estado de interacción en los casos de reinterpretación, y soporta la interpretación incremental de las contribuciones del usuario. La pista de acción de las metas desarrolladas ha sido abordada desde una representación jerárquica que permite estimar las relaciones primarias o secundarias que existen entre las distintas metas de la interacción e identificar, por tanto, el carácter colateral o primario de las contribuciones desarrolladas por los participantes. El Gestor de Metas se incluye para gestionar la inserción, cancelación y resolución de las metas interactivas del sistema. Este componente también tiene asignada la tarea de monitorizar el compromiso establecido sobre las metas combinadas y la criticidad alcanzada sobre las metas individuales del sistema, cuyas variaciones pueden llevar al sistema a tomar, mantener o ceder la palabra, o a producir cualquier otro tipo de contribución en la interacción. Ante los cambios relevantes en el compromiso o en la criticidad de las metas, el Gestor de Metas desencadenará nuevas decisiones de toma de turno.

En lo que respecta a los nuevos componentes del Gestor de Presentación, el Gestor de Continuidad es el componente responsable del tratamiento incremental de los procesos de interpretación y generación de la interacción. Para ello, mantiene una representación parcial de la contribución de los usuarios que va siendo actualizada a medida que recibe de los Componentes de Interfaz nuevos fragmentos de expresión de usuario. Cuando esto sucede, el Gestor de Continuidad combina los nuevos fragmentos de contribución con la contribución parcial que ya tenía con el objetivo de detectar porciones de contribución que, por sí mismas, pudieran representar acciones comunicativas completas (siguiendo la teoría de actos comunicativos). Estos actos comunicativos son enviados al Gestor de Diálogo para que ejecute los progresos oportunos en el estado de interacción y en su conocimiento sociolingüístico asociado. Del lado de la generación, los cambios que van ocurriendo en las circunstancias sociolingüísticas que envuelven a la interacción y el propio estado de interacción motivan que el Gestor de Diálogo, a través de su gestión de metas, desencadene decisiones de toma de turno que serán atendidas en el Gestor de Toma de Turno. Los procesos de decisión de toma de turno pueden desencadenar la formalización, reformulación, rectificación y auto interrupción de las contribuciones de sistema. El gestor de Continuidad es el responsable de combinar estos nuevos fragmentos de contribución con el flujo de contribución en curso del sistema. Durante la gestión de continuidad también deben ser detectados aquellos marcadores de toma de turno que son producidos como alteraciones en la continuidad temporal de la contribución (silencios, silencios oralizados, repeticiones y otros tipos de disfluencia).

Por su parte, la gestión de toma de turno recae sobre el componente Gestor de Toma de Turno, que representa: el estado de actividad los turnos de los participantes (así como la estimación de los posibles TRPs y su carácter primario o secundario); la estimación de qué participante es el considerado hablante actual; y la de cuáles se presentan como candidatos a tomarla (bien por haberla solicitado o por la designación del hablante actual). La actualización del estado de toma de turno es invocada por los componentes Gestor de Continuidad y Gestor de Diálogo cuando son identificados marcadores de toma de turno en la contribución de algún participante, o cuando el estado de interacción o las circunstancias sociolingüísticas cambian durante los procesos de interpretación y generación. El Gestor de Toma de Turno también desarrolla los procesos de decisión de toma de turno, según los cuales decide si contribuir o no en la interacción y determina qué metas puede desarrollar al hacerlo. Para desarrollar estas decisiones, el Gestor de Toma de Turno considera los siguientes conocimientos implicados: el estado de los turnos de los participantes; a quién pertenece la posesión de la palabra; qué participantes son candidatos a tomarla; el compromiso alcanzado sobre las metas interactivas; y la relación beneficio/coste de desarrollarla en la interacción (que depende de las circunstancias

sociolingüísticas en las que se desarrolla la decisión). La decisión de toma de turno comienza con la identificación de aquellas metas que pueden ser desarrolladas por el sistema en la interacción (en ese momento) y termina con una nueva solicitud de generación al Gestor de Diálogo, en la que se adjunta el conjunto de metas que se pueden desarrollar.

Finalmente, el desarrollo simultáneo de los procesos de interpretación y generación hace necesaria una gestión del acceso a los recursos compartidos del sistema (conocimiento relativo al estado de interacción, la sesión, situación, usuarios, emociones, auto modelo) y la regulación del orden con el que esto son ejecutados para garantizar que son cumplidas las restricciones del retardo máximo tolerado con el que los interlocutores esperan recibir la realimentación a sus acciones. El Coordinador de Procesos es el componente que desempeña esta tarea.

Con todo esto, la propuesta de Sistema de Interacción Natural permite el desarrollo de un amplio abanico de posibilidades, en lo que a la formalización de las contribuciones del sistema se refiere, soporta:

- Formalización inicial: El sistema no estaba desarrollando previamente turno y como resultado de la formalización se ha producido una contribución.
- Reformulación: El sistema estaba desarrollando turno, y la expresión de la nueva contribución puede reemplazar la expresión de la previa sin que suponga una ruptura en el flujo de contribución del sistema.
- Rectificación: El sistema estaba desarrollando turno, y la nueva contribución rectifica parte de lo ya expresado en la contribución en curso del sistema.
- Auto interrupción: El sistema estaba desarrollando turno, y la nueva contribución supone una ruptura con la contribución que venía desarrollando hasta el momento (o un fin forzado).

En lo que respecta a las decisiones de toma de turno, estas son abordadas desde una perspectiva de acción combinada que habilitan al sistema para tomar un papel activo en el reparto de los turnos, al igual que hacen los participantes humanos de la interacción. Este planteamiento permite que el sistema aborde una gran diversidad de situaciones de toma de turno:

- Tomar turno para desarrollar una contribución secundaria, independientemente del estado de los turnos, la posesión de la palabra, los candidatos o la urgencia.
- Tomar la palabra si el sistema tiene algo que decir y la palabra está vacante.
- Tomar la palabra si, teniendo el sistema algo que decir, no hay ningún participante desarrollando un turno primario.

- Tratar de tomar la palabra si, teniendo algo que decir y habiendo hablante, éste ha alcanzado un lugar de transición pertinente de la palabra y el sistema se perfila como candidato designado a tomarla (con mayor prioridad si ha sido designado que si es solicitante).
- Solicitar la palabra si, teniendo algo que decir y no estando la palabra vacante, la contribución del hablante no ha alcanzado un TRP pero el sistema considera de cierto grado de urgencia tomar la palabra.
- Interrumpir, aunque exista otro hablante y este no haya alcanzado una posible TRP, si el sistema considera urgente tomar la palabra y las circunstancias sociolingüísticas lo permiten (por ejemplo, por una cuestión de dominancia de su rol frente al del hablante en curso).
- Interrumpir, independientemente de que exista o no hablante y de las circunstancias sociolingüísticas, si el sistema considera de urgencia máxima tomar la palabra (por ejemplo para alertar de peligros).

Del mismo modo, si de lo que se trata es de decidir si el sistema debe continuar o no desarrollando un turno actual, se contemplan los siguientes escenarios:

- Continuar la contribución si, siendo el hablante, existen otros participantes contribuyendo pero no suponen interferencia en su contribución (realizan contribuciones secundarias).
- Continuar la contribución si es secundaria, a pesar de que otros participantes pudieran estar contribuyendo.
- Finalizar su contribución si, siendo el hablante, ha alcanzado una TRP y algún otro participante trata de tomar la palabra (siempre que las circunstancias sociolingüísticas o la urgencia de las metas no determinen lo contrario).
- Reformular su contribución en curso si ocurren cambios en el estado de interacción, las metas interactivas o las circunstancias sociolingüísticas que afecten directamente a la contribución que está desarrollando el sistema, pudiendo resultar en reformulaciones suaves, rectificaciones o auto-interrupciones.

La propuesta ha sido implementada y evaluada sobre la plataforma LaBDA-Interactor. Para la evaluación de la propuesta fueron analizados distintos dominios de interacción en los que se pusieran en juego habilidades avanzadas de toma de turno y en los que la influencia de los problemas asociados a otros niveles de la interacción fueran minimizados (como los relacionados con el reconocimiento y síntesis de voz, el procesamiento de lenguaje natural o

adaptación multimodal). De los dominios analizados, se eligió el dominio de dictado como el mejor candidato, por incluir tanto contribuciones primarias como secundarias, y por desarrollar turnos suficientemente largos como para ser reformulados, solapados e interrumpidos. El dominio incluía diálogos de corrección de errores, solicitudes de ayuda, consultas ortográficas, repeticiones y esperas, entre otros posibles subdiálogos.

Durante la evaluación, se solicitó a los usuarios que interactuaran con distintas configuraciones del sistema con el objetivo de copiar, de forma íntegra y sin errores, un texto dictado por el sistema. Cada una de las configuraciones del sistema abordaba la interacción según una estrategia de toma de turno distinta, siendo las opciones posibles: la toma de turno por ciclo de interacción; la propuesta de toma de turno avanzada; y una toma de turno desarrollada por un participante humano. El usuario no conocía la estrategia desarrollada por el sistema en cada uno de los dictados. Tampoco conocía ni el objetivo de la evaluación, ni el orden en el que cada usuario interactuaba con las distintas configuraciones del sistema.

Durante los experimentos se recogieron de forma automática parámetros técnicos (como el número de metas primarias y secundarias, el porcentaje de metas canceladas y resueltas, el porcentaje de tiempo de posesión de la palabra de cada participante o el número de decisiones de toma de turno desarrolladas por el sistema) y se utilizaron formularios para conocer la satisfacción del usuario y su valoración subjetiva sobre la naturalidad de la interacción (en aspectos como el dinamismo y orden de la interacción, la actitud colaborativa del sistema, lo adecuado de la estrategia de toma de turno para resolver la tarea propuesta, lo cómoda que resultaba la interacción, o la preferencia del usuario de unas configuraciones frente a otras).

Los resultados muestran importantes mejoras de la toma de turno avanzada con respecto al ciclo de interacción, tanto desde un punto de vista técnico como de usabilidad. Del mismo modo, los resultados revelan una elevada naturalidad de esta estrategia de toma de turno en aspectos como la satisfacción del usuario, el orden de la interacción, la actitud colaborativa del sistema, o lo adecuado de la estrategia de toma de turno para el desarrollo de la tarea abordada. La comparación entre la toma de turno avanzada y la toma de turno humana aun muestra margen de mejora para los sistemas de interacción natural, aunque para muchos de los parámetros analizados las diferencias no son significativas.

9.2 *LÍNEAS FUTURAS*

A pesar de que el presente trabajo sienta las bases para el tratamiento de los marcadores de toma de turno producidos como alteraciones de la continuidad temporal de la contribución en curso a través de su componente Gestor de Continuidad (tanto del lado de la interpretación como del de la generación), y a pesar de que algunos de estos marcadores han sido tratados durante su desarrollo (principalmente silencios y silencios oralizados), la gran mayoría de ellos no han podido ser contemplados durante el desarrollo de esta tesis doctoral y su aplicación podría ser de gran utilidad para el desarrollo de estrategias de ganancia, mantenimiento y cesión de la atención (como es el caso de los comienzos y paradas o las paradas y continuación). También han sido tratados algunos tipos de disfluencia en la contribución de los participantes (como las repeticiones, las rectificaciones o los titubeos) pero, en su gran mayoría, tampoco han podido ser abordados. El tratamiento del resto de marcadores de toma de turno y el de las disfluencias sería un objetivo abordable desde las habilidades de Gestión de la Continuidad propuestas con este trabajo. Dada su importancia para un desarrollo natural de la toma de turno, se contempla retomar esta línea de investigación en trabajos futuros. En lo que respecta al resto de marcadores de toma de turno, los que son expresados explícitamente por los participantes de forma verbal o a través del paralenguaje (gestos, miradas, etc.), sólo han sido tratados a nivel de acto comunicativo. Por el momento, su adquisición y síntesis sólo ha sido simulada a través de botones y componentes gráficos en una interfaz de usuario. De este modo, la interpretación y generación de lenguaje natural de estos marcadores no ha sido tratada en el presente trabajo, pero existen estudios que sí abordan este problema y que podrían ser incluidos en versiones futuras del sistema.

Por otro lado, se ha presentado un modelo Gestor de Toma de Turno que permite estimar, tanto el estado de posesión de la palabra como quiénes son los participantes candidatos a tomarla, y aplicar estas conjeturas sobre las decisiones de toma de turno desarrolladas por el sistema (decisiones que se desencadenan por cambios en el estado de la toma de turno y en las metas de la interacción). De esta forma, el sistema puede decidir cuándo debe participar en la interacción, cuando debe continuar haciéndolo o cuando debe parar. Todo ello de una forma flexible a partir de un sistema de toma de turno basado en las reglas descritas por Sacks et al. [149]. Además, el modelo contempla la modulación de estas decisiones en función del estado en el que se encuentran el estado de interacción; las metas propias y combinadas; y las circunstancias que envuelven la interacción. Con ello, el modelo se presenta como una alternativa altamente generalizable e independiente del dominio. Por otro lado, es también un objetivo futuro abordar una separación entre el modelo implementado y el conjunto de reglas

aplicadas a las decisión de toma de turno, con lo que se pretende hacer la propuesta configurable y fácil de adaptar a otros sistemas de toma de turno (distintos del propuesto por Sacks et al.).

La combinación de las mejoras propuestas en este trabajo, en lo que respecta al procesamiento incremental y a la decisión de toma de turno, en combinación con otros modelos de conocimiento no tratados en este trabajo (como los Modelos de Situación, Usuario, Auto Modelo y Ontología) podrían enriquecer la interacción y repercutir de forma muy favorable en la toma de turno. La consideración del Modelo de Situación permitiría aplicar, por ejemplo, aspectos políticos (como la dominancia de unos roles frente a otros) en la decisión de toma de turno. También podría considerarse el aspecto espacial y temporal de la interacción, incluyendo una gestión de eventos que repercutirían directamente sobre la criticidad de las metas propias interactivas y que mejoraría la gestión de situaciones de emergencia (danto, por ejemplo, prioridad a los avisos de riesgos reales para el usuario, como incendios o catástrofes, sobre otros temas de interacción). Del mismo modo, el Modelo de Usuario permitiría adaptar la toma de turno a las propias preferencias de los usuarios, o dar prioridad a unos usuarios frente a otros. También permitiría identificar sus necesidades específicas y revertirlas en la interacción a través de metas propias de recomendación y sugerencias, cuyo desarrollo en la interacción sería gestionado a través de una estrategia de toma de turno avanzada. El Auto Modelo, por representar las metas emocionales del sistema a largo plazo y sus metas operativas, también se beneficiaría de una mejor toma de turno que gestionase los momentos y la forma en la que deben ser desarrolladas en la interacción.

La adaptación multimodal, tanto en lo que a fisión como a fusión se refiere, solo ha sido tratada a nivel superficial a lo largo del presente trabajo. La arquitectura propuesta contempla el lugar que este componente tendría en un Sistema de Interacción Natural completo, los servicios que debería ofrecer a otros componentes y cómo debería comunicarse con todos ellos a efectos prácticos, pero en ningún caso ha sido desarrollado un componente Adaptador Multimodal completo. La interacción humana se desarrolla de forma coordinada a través de las distintas modalidades disponibles en la interacción, las cuales pueden ser distintas en función de los usuarios y de los dispositivos a través de los que acceden al sistema. Gran parte de la comunicación colateral (de la que depende tanto la interpretación combinada que alcanza el hablante a partir de la realimentación simultánea recibida de sus oyentes, como la gestión del reparto de turnos en la interacción) es realizada de forma coordinada a través de distintas modalidades, tal y como demuestran numerosos estudios en el área de la prosodia, la entonación, el timbre y la expresión gestual. Sin un adecuado procesamiento de la adaptación multimodal, esta comunicación colateral no puede ser tratada en la interacción (con la consiguiente pérdida de efectividad de la estrategia de toma de turno desarrollada por el

sistema). Como línea futura, se contempla la integración de los modelos propuestos con un componente de adaptación multimodal que permita sacar el máximo partido a las habilidades de toma de turno del sistema.

Aunque los componentes sobre los que recaen las competencias directas de toma de turno son los ya abordados con este trabajo (Gestor de Continuidad, Coordinador de Procesos; Gestor de Toma de Turno y Gestor de Diálogo), el desarrollo de una toma de turnos avanzada también afecta a los Componentes de Interfaz, tanto a los físicos como a los lógicos, sobre los que un desarrollo de toma de turno avanzada impone una serie de restricciones de obligado cumplimiento. De los componentes físicos de interfaz, tanto de los de entrada como de los de salida, se requiere una granularidad de procesamiento temporal menor al máximo retardo de respuesta que las personas toleran sin perder la percepción de que la interacción se desarrolla en tiempo real (fijada por algunos estudios en torno a 150 y 400 ms.). En esta restricción también se incluye el retardo con el que los Componentes de Interfaz desarrollan la adquisición y síntesis de las expresiones naturales. Del mismo modo, los componentes físicos de salida deben soportar la confirmación de los fragmentos de expresión de las contribuciones del sistema y permitir la cancelación de la expresión en curso. La recepción de confirmaciones de expresión de los fragmentos de contribución es requerida por el Gestor de Continuidad para determinar cuándo pueden darse por expresados cada uno de los actos comunicativos individuales que componen la contribución del sistema, requisito indispensable para un correcto procesamiento incremental de la interacción. Respecto a la cancelación de la expresión en curso, se requiere de los Componentes de Interfaz la capacidad para interrumpir la expresión en curso del sistema, contemplándose la capacidad de confirmar anticipadamente aquellos fragmentos de expresión cuya síntesis esté prevista durante el intervalo de tiempo que se requeriría para tramitar una hipotética reformulación de la contribución en curso, con el fin de garantizar la mayor fluidez posible en la contribución producida por el sistema y minimizar sus discontinuidades.

En lo que respecta a los Componentes de Interfaz Lógicos, su operación estará coordinada por el Gestor de Continuidad (componente que integra fragmentos de contribución en contribuciones completas), y las decisiones de toma de turno se verán favorecidas en la que medida en que esos componentes sean capaces de emitir proyecciones anticipadas de los actos comunicativos en curso que se reciben a la entrada, y de estimar aquellos puntos de las expresiones de salida en los que sus interlocutores pudieran alcanzar proyecciones anticipadas de los actos comunicativos del sistema. Esta gestión de las proyecciones anticipadas permitiría obtener predicciones tempranas de los lugares de transición pertinentes de la palabra (al igual que hacen los participantes humanos de la interacción) y alcanzar con ellas un mejor compromiso en torno al estado en el que se encuentra la toma de turno. La proyección

anticipada de actos comunicativos mejora el aprovechamiento de tiempo en la interacción, reduciendo los silencios existentes durante los traspasos la palabra y mejorando las opciones de los candidatos a tomar la palabra para erigirse en hablantes siguientes.

Como línea de trabajo futura, también se contempla la aplicación del sistema implementado a otros dominios de interacción que pudieran beneficiarse de una estrategia de toma de turno avanzada. Entre estos dominios podrían encontrarse la navegación, el entretenimiento, la educación y la asistencia a usuarios, aunque en términos generales, cualquier dominio en el que se requiera un alto grado de pro actividad del sistema, o la consideración de unas circunstancias sociolingüísticas cambiantes, se vería altamente beneficiado por las nuevas habilidades del sistema.

GLOSARIO

Acción: Ejecución de un efecto sobre un entorno, a resultas de un proceso cognitivo [21]. Puede ser:

· Individual: La ejercida por un único participante.

· Autónoma: Que no precisa de acciones complementarias de otros agentes.

· Participatoria: Que no es completa en sí misma, precisando una acción complementaria.

· Combinada: Ejercida por varios agentes de forma colaborativa. Acción que sólo resulta completa de la combinación de las acciones individuales y participatorias de todos los agentes.

Actividad: Acción orientada a la consecución de una *meta* o conjunto de metas [21].

Acto Comunicativo: Generalización del concepto de *acto de habla* a todas las posibles modalidades de la interacción.

Acto de Habla: Cada uno de los actos materializados a través del habla que designan acciones y se realizan al designarlas. Cualquier acto de habla conlleva actos a los niveles locutivos (aquello que se dice), ilocutivo (la intención o finalidad concreta de lo que se dice) y perlocutivo (los efectos que se pretenden producir en el interlocutor al enunciarlos). Los actos de habla en los que el acto locutivo y el ilocutivo coinciden constituyen actos directos, mientras que aquellos en los que la intención comunicativa no se indica de forma intencionada y explicita se denominan actos indirectos. En función de su intencionalidad y finalidad pueden ser asertivos o expositivos, directivos, compromisorios, declarativos y expresivos [5;159].

Adaptación Multimodal: Conversión de un flujo de datos multimodal a un flujo de datos monomodal y viceversa, que comprende tanto a los procesos de *fusión* como a los de *fisión* multimodal.

Agente Interactivo: Entidad autónoma capaz de participar en la interacción con el objetivo de satisfacer sus propias *metas* y que acepta que, para conseguirlo, debe colaborar en la resolución de las metas del resto de agentes interactivos. Los agentes interactivos pueden ser tanto humanos como máquina.

Asistentes Virtuales: *Véase Sistema de Interacción Natural*.

Auto interrupción: Caso de *reformulación* en la que se determina que el participante no debe generar nada o que debe generar una *contribución* que supone una ruptura temática con la contribución en curso que viene desarrollando el participante.

Circunstancias Sociolingüísticas: Conjunto de realidades que caracterizan al entorno social y lingüístico en el que se desarrolla la interacción. Esta información condiciona y determina el desarrollo de la propia interacción. Incluye el conocimiento representado en los Modelos de Sesión, Usuarios, Situación y Emocional, y también en la propia Ontología.

Compromiso: Acuerdo establecido entre dos o más agentes para conseguir algo (una *meta* conjunta persistente) en una *zona común* y en unas *circunstancias sociolingüísticas* dadas, partiendo del hecho de que todos los implicados conocen la meta y el compromiso, que todos creen que los demás han aceptado el compromiso, y que todos creen que esto se aplica a todos los implicados [21].

Confirmación de Expresión: Subproceso de la *generación*. Consiste en la actualización del estado de interacción con los movimientos asociados a la expresión de los actos comunicativos de la contribución en curso del sistema. Dicho proceso de ejecuta cada vez que pueden darse por sintetizadas todas las expresiones que componen la enunciación de alguno de los actos comunicativos de la contribución.

Conocimiento Sociolingüístico Asociado: *Véase Circunstancias Sociolingüísticas*.

Contexto: Información lingüística que caracteriza a una *sesión* y que será útil al progreso de la interacción. Se trata de información estática no permanente, como por ejemplo la historia de la conversación.

Contribución Primaria: Cada una de las *contribuciones* que producen movimientos en los asuntos oficiales de la interacción y que hacen progresar las *metas* que la motivan. Por lo general, los participantes sólo producen estas contribuciones primarias cuando se encuentran en posesión de la *palabra*, aunque existen situaciones en las que podrían colisionar contribuciones primarias de distintos participantes. Esto ocurre con frecuencia durante los traspasos de *palabra*, en los que el comienzo de la contribución primaria del *hablante* siguiente podría solaparse con el final de la contribución actual del hablante en curso. Esta estrategia también puede ser aplicada para forzar el cambio de hablante durante los *TRPs* de la contribución primaria del hablante en curso. Del mismo modo, algunas contribuciones primarias no requieren estar en posesión de la palabra para ser expresadas (por ejemplo aquellas que pueden ser expresadas a través de modalidades alternativas a la de la contribución primaria del hablante en curso). No obstante, la producción simultánea de contribuciones primarias de distintos participantes tiende a desencadenar luchas por la posesión de la palabra.

Contribución Secundaria: *Contribuciones* de carácter metacomunicativo que se expresan con el objetivo de mejorar la calidad de la comunicación y mostrar evidencias de cierre a las acciones del *hablante* en curso. Las contribuciones secundarias suelen caracterizarse por ser cortas y poder ser expresadas a través de modalidades alternativas al habla (como gestos o miradas). No hacen progresar por sí mismas los asuntos oficiales de la interacción, aunque son necesarias para gestionarla. Algunos ejemplos de contribuciones secundarias son las solicitudes y cesiones de palabra, la realimentación simultánea, risas o carraspeos.

Contribución: Conjunto de expresiones que un participante desarrolla (o pretende desarrollar) en la interacción a lo largo de un mismo *turno*. Conlleva una secuencia de *actos comunicativos* a realizar por parte del participante. La producción de una contribución conlleva una formalización inicial previa y un posterior desarrollo de su expresión, a lo largo del cual la contribución podría ser reformulada (o incluso interrumpida). Una contribución pretende hacer progresar la interacción a los niveles local y global.

Conversación: Acción y efecto de interactuar entre varios agentes según procedimientos, modalidades y códigos naturales a las personas, bien sea de forma estructurada o espontánea, que se desarrolla en un contexto definido y que puede girar en torno a diversos temas.

Corpus: *1.* Conjunto ordenado de *diálogos* que son recogidos a partir de las interacciones realizadas por participantes en los *escenarios* descritos en un *dominio de interacción* dado. En el corpus los diálogos pueden recogerse en forma de transcripciones y grabaciones, y pueden contener diversos niveles de anotación. *2.* Estructuras de conocimiento y

formalizaciones aplicadas sobre un *Sistema de Interacción Natural* que permiten adaptarle a un dominio de interacción específico y le capacitan para desarrollar interacciones en él.

Criticidad: Función propia de cada una de las *metas individuales* de un participante que determina la importancia que tiene para dicho participante hacer progresar la *meta. E*s dependiente del *estado de interacción*, de las *circunstancias sociolingüísticas* y del *estado de toma de turno*.

Decisión de Toma de Turno: Proceso por el cual el participante determina si debe o no contribuir en la interacción o si, en caso de estar actualmente desarrollando un *turno*, debe o no continuar haciéndolo. Las decisiones de toma de turno se desencadenan ante los cambios producidos en el *estado de toma de turno*, ante los cambios producidos en el *compromiso* de las metas compartidas por todos los participantes o en la *criticidad* de sus metas individuales. Las decisiones de toma de turno limitan el conjunto de *metas* que el participante puede desarrollar en la interacción y son dependientes del momento en el que se ejecuta la decisión. La decisión de toma de turno se realiza a partir del cálculo de la *urgencia* de desarrollar cada una de las metas y de lo *favorable* que sea el estado de posesión de la *palabra* a la toma de la palabra del participante.

Diálogo: *1.* Cada una de las realizaciones de uno o varios *escenarios* que los participantes desarrollan en una interacción real y de acuerdo a un *dominio de interacción* dado. *2.* Interacción desarrollada entre dos participantes. *3.* Fragmento de sesión con sentido intencional independiente.

Dominio de Interacción: Ámbito al que se restringe la interacción que pueden mantener los participantes. El dominio de interacción determina el conjunto de habilidades, tanto interactivas como operativas, que cada participante podrá desarrollar en la interacción, así como los *escenarios* posibles en los que estas serán puestas en juego.

Escenario: Cada una de las posibles situaciones de interacción que pueden desarrollarse dentro de un *dominio de interacción* dado, en las que se pondrán en juego un conjunto de habilidades, tanto interactivas como operativas, de cada uno de los participantes. Los escenarios pueden mantener entre sí una relación jerárquica, de más generales a más específicos, y ser desarrollados de forma combinada en la interacción.

Estado de Interacción: Situación de progreso en la que se encuentra cada uno de los *hilos* de la interacción y su contexto asociado, las relaciones jerárquicas establecidas entre ellos en la

estructura intencional y el orden actual de atención prestado a los hilos en la *estructura focal*.

Estado de Toma de Turno: Conocimiento relativo a la situación actual en la que se encuentra la interacción a un nivel de *organización temporal*. La representación del estado de toma de turno se sustenta en conjeturas que realiza el participante sobre el estado en el que se encuentra el turno de cada participante (incluido el suyo propio), quién es el participante que ostenta la posesión de la *palabra* y quiénes son los candidatos a tomarla.

Estructura Focal: Lista de las instancias de *hilo* abiertas en la interacción que se encuentran ordenadas según la atención que los participantes les han prestado históricamente. El primero de los hilos de la estructura focal es el *foco*.

Estructura Intencional: Representación de la relación jerárquica existente entre las distintas instancias de hilo abiertas en la interacción. La raíz es una instancia del denominado hilo base, que representa el desarrollo de la propia interacción.

Expresión: Unidad mínima interpretable o generable en una interacción.

Favorable: Dícese de la posesión de la palabra con respecto a un participante cuando el *estado de toma de turno* hace adecuado que dicho participante tome la *palabra* en la interacción para desarrollar cualquier tipo de *meta* (sin restricción). La posesión de la palabra será favorable a un participante cuando la palabra esté vacante, cuando éste ya ostente la posesión de la palabra, o cuando sea candidato a tomarla y se den determinadas condiciones relativas al estado del turno del hablante actual (por ejemplo, haber ocurrido una *TRP*).

Fisión Multimodal: Proceso por el cual se seleccionan las modalidades más adecuadas para la síntesis de la *contribución* generada por el sistema y se descompone de forma coordinada dicha contribución en las expresiones correspondientes a esas modalidades.

Foco: Puntero a la instancia de *hilo* que desarrolla la meta combinada sobre la que se centra la atención de los participantes en un momento dado. Primera instancia de hilo de la *estructura focal*.

Formalización: Fase inicial del desarrollo de la *generación*. Durante la formalización el participante diseña la *contribución* que pretende producir en la interacción y determina los movimientos que la expresión de cada uno de los *actos comunicativos* que la componen producirá en el *estado de interacción*.

Fragmento de contribución con sentido interactivo completo: Se considerarán fragmentos de contribución con sentido interactivo completo a cada porción, aun incompleta, de una *contribución* que puede ser interpretada o generada por si misma a nivel de diálogo, produciendo movimientos en el estado de la interacción. Se relacionan con aquellas porciones de contribución cuyo procesamiento de lenguaje natural produce como resultado un acto comunicativo completo, por ser estas las unidades mínimas que producen movimientos en el estado de interacción.

Fragmento de contribución: Desde un punto de vista incremental de la interacción, las *contribuciones* de los participantes van siendo desarrolladas poco a poco en el tiempo. De esta forma, las contribuciones no son tratadas por el sistema como unidades atómicas e indivisibles, sino que van siendo interpretadas y generadas poco a poco, en unidades menores. Estas unidades menores son denominadas en este trabajo fragmentos de contribución. La granularidad de los fragmentos de contribución puede variar en función de las necesidades de los participantes, y suele ser frecuente procesar las contribuciones a nivel de palabra o en intervalos de tiempo de duración determinada.

Fusión Multimodal: Proceso por el cual un sistema es capaz de combinar los flujos de información recibidos desde distintos componentes de adquisición (micrófonos, cámaras, ratones, teclados) a un flujo de información único que podrá ser procesado de forma homogénea por el sistema.

Generación: Proceso según el cual el sistema produce sus contribuciones en la interacción. Desde un enfoque incremental de la interacción, la generación se descompone en las fases de *formalización* inicial de la contribución (que define los movimientos que irá produciendo la contribución sobre el estado de interacción durante su desarrollo) y la *confirmación de expresión* de los actos comunicativos que la componen (que ejecuta sobre el estado de interacción los movimientos que se asocian a la expresión de dichos actos comunicativos).

Hablante: Participante sobre el que recae la posesión de la palabra en la interacción en un momento dado. La estimación del hablante se basa en conjeturas realizadas por todos los participantes de la interacción de forma combinada, y en ella se considera: qué participantes desarrollan *turno*; que participantes son candidatos a tomar la *palabra*; cuál es el *estado de la interacción* y el de las *metas* interactivas; y cuáles son las *circunstancias sociolingüísticas* en las que se desarrolla la interacción. En general, será considerado hablante aquel participante que se encuentre desarrollando una contribución primaria en la interacción, aunque podría no encontrarse desarrollándola en exclusiva, o no desarrollarla aun recayendo sobre él la palabra.

Hilo: Formalización de los posibles desarrollos de una *meta* en la interacción. Representa cada uno de los posibles movimientos que pueden producirse en el estado de progreso de la meta que representa y que son producidos como consecuencia de la expresión de *actos comunicativos* por parte de los participantes. El hilo formaliza las tareas que debe realizar el sistema en la interacción en cada uno de sus posibles estados y qué actos comunicativos debe generar para hacer progresar la meta. Un hilo es instanciado en la interacción cuando alguno de los participantes introduce la necesidad de resolver su meta asociada y el desarrollo del hilo en la interacción constituye un *segmento* en el *diálogo*.

Intención: Pretensión de algún efecto o acción futura [21].

Interacción Hombre Máquina: Disciplina que estudia el intercambio de información que se produce entre las personas y los ordenadores y que se encarga del diseño, evaluación e implementación de los aparatos tecnológicos interactivos.

Interacción Humana: Interacción desarrollada entre personas según lenguas y procedimientos naturales a los participantes,. En ella se contempla tanto una comunicación verbal como no verbal a través de cualquiera de las modalidades disponibles (habla, gestos, miradas, etc.).

Interacción Natural: Interacción desarrollada entre personas y un Sistema de Interacción Natural desarrollada según modalidades, códigos y procedimientos naturales a los participantes humanos de la interacción. En este tipo de interacción, el mayor esfuerzo de adaptación a las necesidades de su interlocutor debe recaer del lado del sistema. El objetivo último de este paradigma de Interacción Hombre Máquina sería alcanzar una interacción idéntica a la desarrollada con un participante humano.

Interlocutor: Para un participante dado, cada uno de los agentes, tanto humanos como máquinas, que participan junto a él en el desarrollo de la interacción y la hacen progresar a través de las contribuciones que expresan en forma de turnos, bien sean primarios o secundarios.

Interpretación: Proceso según el cual un participante comprende el sentido de la contribución de otro participante y actualiza con él su *estado de la interacción* y el *conocimiento sociolingüístico asociado*.

Interrupción: Cese forzado de la *contribución* de un participante, producida como consecuencia de cambios producidos en las *circunstancias sociolingüísticas* o motivado por una colisión con las contribuciones de otros participantes.

Intervalo: *Silencio* producido por los participantes como recurso para ceder la *palabra* a otros participantes. Por tanto, los intervalos ocurren durante los *TRPs* y se producen entre el fin de la *intervención* del *hablante* en curso y el comienzo de la del hablante siguiente.

Intervención: *Véase contribución primaria.*

Lapso: *Silencios* ejecutados por los participantes durante su *contribución* para denotar un cambio de *segmento*.

Lugar de Transición Pertinente (TRP): Concepto introducido por Sacks et al. [149], del inglés Transition Relevance Place, que determina aquellos puntos de la intervención de un participante en los que podría producirse un traspaso de la *palabra*. Estos puntos pueden coincidir con el final de un *turno*, aunque no tienen por qué hacerlo necesariamente, y tras ellos el hablante podría expresar nuevas *unidades de construcción del turno*, o TCUs (del inglés Turn Construction Unit).

Meta discursiva propia: *Véase Meta Individual*.

Meta Compartida: Objetivo común comprometido por varios participantes para satisfacer sus propias *metas individuales*. Su resolución requiere de las acciones participatorias de todos los participantes.

Meta Individual: Objetivo propio e individual de un participante a cuya resolución condiciona sus acciones y que motiva su participación en la interacción.

Meta: Objetivo que rige las acciones de cualquier agente en la interacción.

Modalidad: Cada uno de los posibles canales de comunicación en que los participantes pueden realizar sus expresiones. La modalidad dominante en la interacción humana es el habla, aunque existen otras modalidades de gran importancia, como son la gestual, las miradas, los movimientos corporales, etc.

N-grama: Subsecuencia de n elementos, por ejemplo palabras, de una secuencia mayor. Bigrama y trigrama consisten en n-gramas de tamaños dos y tres, respectivamente.

Organización Temporal de la Interacción: *Véase Toma de Turno.*

Oyente: Aquellos participantes que reciben la *contribución* primaria del *hablante* y que colaboran en su construcción a través de las evidencias de cierre que le ofrecen en forma de

realimentación simultánea a su contribución (según el principio de la interpretación combinada [77]).

Palabra: Derecho para realizar *contribuciones primarias* en la interacción. La palabra sólo puede ser ostentada por un participante en cada momento, según la regla del hablante único [32], y su traspaso entre los distintos participantes es regulado como una acción combinada entre ellos.

Participante: Cada uno de los agentes, tanto humanos como máquinas, que tienen intereses comunes en el desarrollo de la interacción y la hacen progresar a través de las contribuciones que expresan en forma de *turnos*, bien sean primarios o secundarios.

Pausa: *Silencio* que ocurre en el seno de una *intervención* y que es producido por el *hablante* como recurso de ganancia de la atención de sus *interlocutores*.

Presuposición: Conocimiento compartido por los participantes en una interacción que no proviene de las expresiones ejercidas en la misma, sino que es intrínseco a las *circunstancias sociolingüísticas* en las que se desarrolla.

Pro actividad: Acción de insertar nuevas *metas* en la interacción por iniciativa propia.

Pro activo: Dícese del *agente interactivo* que es capaz de insertar nuevas *metas* en la interacción por iniciativa propia.

Rectificación: Caso de *reformulación* en la que la nueva *contribución* formalizada supone una corrección de parte de lo ya expresado en la contribución en curso del participante, por lo que la generación debe retomar parte de lo ya expresado y aplicar estrategias encaminadas a corregir y reorientar el curso de la contribución.

Reformulación suave: Caso de *reformulación* en la que la nueva *contribución* formalizada coincide con la contribución en curso del participante (o no afecta a lo ya expresado) y puede continuarse con la *generación* de forma fluida.

Reformulación: Caso especial de *formalización* que se desencadena durante la producción de otra *contribución* del mismo participante y que tiene como objetivo determinar si sigue siendo pertinente continuar con la *generación* de dicha contribución o sí, por el contrario, ésta debe ser revisada o interrumpida. En función de sus posibles resultados puede ser una *reformulación suave*, una *rectificación* o una *auto interrupción*.

Robots Conversacionales (Chatbots): *Véase Sistema de Interacción Natural*.

Segmento: Porción de *diálogo* que desarrolla una *meta compartida* [21].

Sesión: Intervalo de tiempo en el que se desarrolla una interacción completa [21].

Silencio: Inactividad producida durante el desarrollo del *turno* de un participante [142, pp.126]. Los silencios se caracterizan por su duración y caída tonal, pero su correcta *interpretación* depende del *estado de la toma de turno*, del *estado de interacción* y de las *circunstancias sociolingüísticas* en las que se produce, por lo que su categorización se realiza durante la gestión del diálogo. Los silencios pueden clasificarse entre *pausas*, *intervalos* y *lapsos*.

Sistema de Diálogo: *Véase Sistema de Interacción Natural*.

Sistemas de Interacción Natural: Cualquiera de los sistemas que aspiran a hacer accesible la tecnología a través de una interacción basada en los mismos procedimientos, códigos y modos que las personas aplican de forma natural en su interacción humana. Alcanzar de forma completa es, hoy por hoy, una utopía y para muchos autores se trata en realidad de Sistemas de Interacción Quasi-Natural (Natural-Like Interaction Systems) que simulan el lenguaje natural en un dominio concreto, con un histórico pequeño y con estructuras de lenguaje especializadas. La literatura se ha venido comúnmente refiriendo a estos sistemas como Sistemas de Diálogo, Sistemas de Interacción por Voz (Spoken Dialogue Systems), Asistentes Virtuales o Robots Conversacionales (Chatbots).

Sistemas de Interacción por Voz (Spoken Dialogue Systems): *Véase Sistema de Interacción Natural*.

Solapamiento: Situación de contribución simultánea entre distintos participantes, bien sea el producido entre *contribuciones primarias*, *secundarias* o la combinación de ambas.

Subdiálogo: Fragmento de un diálogo que tiene su propia línea de discurso y, por tanto, desde la interpretación intencional es (parcialmente) independiente del resto del diálogo [21].

Toma de Turno Avanzada: *Véase Toma de Turno Natural.*

Toma de Turno Humana: La desarrollada por participantes humanos en su interacción.

Toma de Turno Natural: La desarrollada en interacciones entre humanos y máquinas que pretende ser desarrollada según los mismos procedimientos, códigos y modalidades que la *toma de turno humana*.

Toma de Turno: Actividad que organiza los intervalos de contribución de los participantes en la interacción y permite que todos ellos tengan la posibilidad de ser tanto emisores como receptores de los intercambios comunicativos. Determina tanto los momentos en los que cada participante debe contribuir, como la forma en la que todos ellos organizan su contribución a lo largo del tiempo.

Turno: Acción de ocupar el canal en la interacción durante un periodo de tiempo con el objetivo de expresar una *contribución*. El termino turno hace referencia a la acción realizar participaciones en la interacción, más allá de los contenidos que son desarrollados en él y de los progresos que su contribución supone en el estado de desarrollo de las *metas* o en las *circunstancias sociolingüísticas* de la interacción.

Unidad de Construcción del Turno (TCU): Una Unidad de Contrucción del Turno, o TCU (del inglés Turn Construction Unit), es el segmento fundamental de habla en la conversación, tal y como se describe desde el análisis conversacional introducido por Sacks et al. [149]. Describe cada una de las piezas de la conversación que pueden ser interpretadas por los participantes como un turno completo en la interacción (independientemente de que esa fuera o no la intención del participante que la expresa). Los finales de las TCUs son los denominados *lugares de transición pertinentes*, o TRPs (del inglés Transition Relevance Place).

Urgencia: Resultado de la evaluación de la función de *criticidad* de una *meta individual* y del *compromiso* asociado a su *meta combinada* bajo unas determinadas condiciones del *estado de interacción*, las *circunstancias sociolingüísticas* y el *estado de toma de turno*. Esta urgencia puede ser muy alta, alta, media o baja en función de que la *meta*: deba ser desarrollada de inmediato; deba ser desarrollada de inmediato solo si se cumplen determinadas condiciones; deba ser señalizado que se desea desarrollarla; o de si el participante puede esperar a estar en posesión de la *palabra* para desarrollarla.

Zona Común: Plano de interpretación de las *metas compartidas* durante la interacción [21].

BIBLIOGRAFÍA

1. Allen, JF. (1996). *Logical form in the trains-96 system*. Technical Report, University of Rochester, 1996.

2. Allen, JF. Byron, D., Dzikovska, M., Ferguson, G., Galescu, L., Stent, A. (2001). *Towards Conversational Human-Computer Interaction*. AI Magazine.

3. Andry, F., Bilange, E., Charpentier, F., Choukri, K., Ponamalé, M., and Soudoplatoff, S. (1990). *Computerised simulation tools for the design of an oral dialogue system*. In Selected Publications, 1988-1990, SUNDIAL Project (Esprit P2218). Commission of the European Communities.

4. Atterer, M., Baumann, T., Schlangen, D. 2008. *Towards incremental end-of-utterance detection in dialogue systems*. In: Coling, Manchester, UK. pp. 11–14.

5. Austin, JL. (1962). *How to Do Things With Words*. Oxford University Press, Oxford.

6. Bellifemine, F., Caire, G., Poggi, A., Rimassa, G. (2003). *JADE - A White Paper*. Exp. Vol.3, nº3.

7. Beringer, N., Kartal, U., Louka, K., Schiel, F., Türk, U. (2002). *PROMISE: A procedure for multimodal interactive system evaluation*. In Proceedings of the LREC Workshop on Multimodal Resources and Multimodal Systems Evaluation, pp. 77–80. Las Palmas, Spain.

8. Bernsen, N. (2002). *Speech-related technologies: where will the field go in 10 years?* Association for Computational Linguistics, Vol. 13, pp.1-19. Morristown, NJ, USA.

9. Bernsen, N.O., Dybkjær, L. (2000). *A methodology for evaluating spoken language dialogue systems and their components*. In Proceedings of the Second International Conference on Language Resources and Evaluation, pp. 183–188. Athens.

10. Bertini, E., Santucci, G. (2004). *Modelling internet based applications for designing multi–device adaptative interfaces*. In Working conference on advanced visual interfaces, pp.252–256.

11. Beskow, J. McGlashan, S. (1997). *Olga: A Conversational Agent with Gestures*. In Proceedings of the IJCAI'97 workshop on Animated Interface Agents - Making them Intelligent. Morgan-Kaufmann Publishers, San Francisco.

12. Besser, J., Alexandersson, J., (2008). *A comprehensive disuency model for multi-party interaction*. Proceedings of the 8st SIGdial Workshop on Discourse and Dialogue.

13. Black, WJ., Bunt, HC., Dols, FJH., Donzella, C. Ferrari, G., Haidan, R., Imlah, WG., Jokinen, K., Lager, T., Lancel, JM., Nivre, J., Sabah, G., Wachtel, T. (1991). *A Pragmatics-Based Language Understanding System*. PLUS Deliverable D1.2.

14. Bobrow, DG., Kaplan, RM., Kay, M., Norman, DA., Thompson, H., Winograd, T. (1977).*GUS: A frame-driven dialog system*. Artificial Intelligence, Vol.8, Issue.2, pp.155-173. Elsevier Science B.V.

15. Boomer, DS. 1965. Hesitation and grammatical encoding. Language and Speech, Vol.8, pp.148-158.

16. Braubach, L., Pokahr, A., Lamersdorf, W. (2005). *Jadex: A BDI-Agent System Combining Middleware and Reasoning*. Software Agent-Based Applications, Platforms and Development Kits. Whitestein Series in Software Agent Technologies and Autonomic Computing, pp.143-168.

17. Brown, P. Levinson, SC. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

18. Brown, SM., Santos EJr., Banks, SB. (1999). *Active user interfaces for building decision-theoretic systems*. In Proceedings of the 1st Asia-Pacific Conference on Intelligent Agent Technology, pp.244-253. Hong Kong.

19. Bruce, B. (1981). *A social interaction model of reading*. Discourse Processes, Vol.4, pp.273-311.

20. Calle-Gómez, F.J., García-Serrano, A., Martínez-Fernández, P. (2006). *Intentional processing as a key for rational behaviour through Natural Interaction*. Interacting with Computers. Vol. 18, pp. 1419-1446. ELSEVIER.

21. Calle-Gómez, FJ. (2004). *Interacción Natural Mediante Procesamiento Intencional: Modelo de Hilos en diálogos*. Ph.D. dissertation (Spanish), Univ. Politécnica de Madrid, Spain, 2004.

22. Cañadas-Osinski, I., Sánchez-Bruno, A. (1998). *Categorías de respuesta en escalas tipo Likert*. Psicothema, Vol.10, nº3, pp. 623-631. Oviedo, Spain.

23. Carretero, MP., Oyarzun, D., Aizpurua, I., Ortiz, A. (2004). *Animación Facial y Corporal de Avatares 3D a partir de la edición e interpretación de lenguajes de marcas*. CEIG.

24. Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson H., Yan, H. (1999). *Embodiment in Conversational Interfaces*. Rea. CHI'99, Pittsburgh, PA.

25. Cathcart, N., Carletta, J., Klein, E. 2003. *A shallow model of backchannel continuers in spoken dialogue*. In: EACL, pp.51–58.

26. Chafe, W. (1979). *The flow of thought and the flow of language*. In T. Givon (Ed.), Syntax and semantics I2: Discourse and syntax, pp.159-181. Academic Press., New York.

27. Chafe, W. (1980). *The deployment of consciousness in the production of narrative*. In W. Chafe (Ed.), The pear stories, pp.9-50. Norwood NJ: Ablex.

28. Chafe, W. (1992). *Intonation units and prominences in English natural discourse*. Paper presented at the University of Pennsylvania Prosodic Workshop, Philadelphia.

29. Cheyer, A., Martin, D. (2001). *The Open Agent Architecture*. Journal of Autonomous Agents and Multi-Agent Systems, Vol.4, Issue. ½, pp.143–148.

30. Churcher, GE, Atwell, ES, Souter, CA. (1997). *Generic Template To Evaluate Integrated Components In Spoken Dialogue Systems*. Workshop On Interactive Spoken Dialog Systems: Bringing Speech And NLP Together In Real Applications, pp. 9-16.

31. Cicourel, A. (1992). *The interpenetration of communicative contexts: examples from medical encounters*. In A. Duranti and C. Goodwin (Eds.) Rethinking Context: Language as an Interactive Phenomenon, p.291-310. Cambridge University Press., Cambridge.

32. Clark, HH. (1996). *Using Language*. Cambridge University Press.

33. Clark, HH., Schaefer, EF. (1987). *Collaborating on contributions to conversation*. Language and Cognitive Processes, Vol.2, pp.19-41.

34. Clark, HH., Schaefer, EF. (1989). *Contributing to discourse*. Cognitive Science, Vol.13, pp.259–294.

35. Clark, RAJ., Richmond, K., King. S. (2007). *Multisyn: Open-domain unit selection for the Festival speech synthesis system*. Speech Communication, Vol.49, Issue.4, pp.317-330.

36. Cohen, PR. (1997). *Dialogue Modeling*. In Survey of the state of the art in Human Language Technology, chap.6, pp.204-209. Cambridge Univ. Press.

37. Colby, KM., Weber, S., Hilf, FD. (1971). *Artificial paranoia.* Artificial Intelligence, Vol.2, pp.1-25.

38. Cole, R. (1997). *Survey of the State of the Art in Human Language Technology*. In R. Cole, J. Mariani, H. Uszkoreit, G. Batista-Varile, A. Zaenen, A. Zampolli, V. Zue (Eds.), Cambridge Uni. Press and Giardini, pp. 234-240.

39. Cooper, FS. (1953). *Some Instrumental Aids to Research on Speech*, Report on the Fourth Annual Round Table Meeting on Linguistics and Language Teaching , pp.46-53. Georgetown University Press.

40. Cuadra-Fernández, D., Calle-Gómez, FJ., Rivero-Espinosa, J., Valle-Agudo, D. (2008). *Applying Spatio-Temporal Databases to Interaction Agents*. International Symposium on Distributed Computing and Artificial Intelligence, Advances in Soft Computing, Vol. 50, pp.536-540. Springer Berlin / Heidelberg.

41. Cuadra-Fernández, D., Rivero-Espinosa, J., Valle-Agudo, D., Calle-Gómez, FJ. (2008). *Enhancing Natural Interaction with Circumstantial Knowledge*. In Richards, D. (Ed.) International Transactions on Systems Science and Applications, Vol. 4, Issue.2, pp.122-129. Springer, Heidelberg.

42. Cutler, EA., Pearson, M. 1986. On the analysis of prosodic turn-taking cues. In: Johns-Lewis, C. (Ed.), Intonation in Discourse. College-Hill, San Diego,CA. pp.139–156.

43. Dudley, H., Riesz, RR., Watkins, SSA. (1939). *A synthetic speaker*. J. Franklin Inst, Vol. 227, nº.6, pp.739-764.

44. Duncan, S. (1973). *Toward a grammar for dyadic conversations*. Semiotica, Vol.9, pp.29-46.

45. Duncan, S. 1972. *Some signals and rules for taking speaking turns in conversations*. Journal of Personality and Social Psychology. Vol.23, pp.283–292.

46. Duncan, S. 1974. *On the structure of speaker-auditor interaction during speaking turns*. Language in Society. Vol.3, pp.161–180.

47. Duncan, S. 1975. *Interaction units during speaking turns in dyadic, face-to-face conversations*. In Organization of Behavior in Face-to-Face Interaction. Mouton Publishers, Den Hague, pp.199–213.

48. Duncan, S., Fiske, D. 1977. *Face-To-Face Interaction: Research, Methods, and Theory*. Lawrence Erlbaum Associates.

49. Duranti, A., Brenneis, D. (1986). *The audience as co-author*. Special issue of Text, Vol.6, Issue.3, pp. 239-47.

50. Duranti, A., Goodwin, C. (1992). *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge University Press., Cambridge.

51. Dybkjær, L., Bernsen, NO. and Minker, W. (2004). *Evaluation and Usability of Multimodal Spoken Language Dialogue Systems*. Speech Communication, Vol.43, pp.33-54.

52. Edlund, J., Heldner, M., Gustafson, J. 2005. *Utterance segmentation and turn-taking in spoken dialogue systems*. Sprachtechnologie mobile Kommunikation und linguistische Ressourcen, pp.576–587.

53. Eggins, S., Slade, D. (1997). *Analysing casual conversation*. Cassell, London.

54. Ekman, P., Friesen, WV. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. Prentice Hall, New Jersey.

55. Ertl, MA., Gregg, D., Krall, A., Paysan, B. (2002). *VMGEN: A generator of efficient virtual machine interpreters*. Software: Practice and Experience, Vol.32 Issue.3, pp.265-294.

56. Fernández Martínez, F. (2008). *Análisis, diseño y aplicación de modelos de diálogo flexibles, contextuales y dinámicos basados en redes bayesianas*. Tesis Doctoral, E.T.S.I. Telecomunicación (UPM).

57. Ferrer, L., Shriberg, E., Stolcke, A. 2002. *Is the speaker done yet? Faster and more accurate end-of-utterance detection using prosody*. In: Proceedings of theICSLP, pp.2061–2064.

58. Ferrer, L., Shriberg, E., Stolcke, A. 2003. *A prosody-based approach to end-of-utterance detection that does not require speech recognition*. In: Proceedings ofICASSP.

59. Ford, C., Thompson,S. 1996. *Interactional units in conversation: syntactic, intonational and pragmatic resources for the management of turns*. In: Ochs, E., Schegloff, E., Thompson, S. (Eds.), Interaction and Grammar. Cambridge University Press, pp.134–184.

60. Ford, CE., Thompson, SA. (1996). *Interactional Units in Conversation: Syntactic, Intonational, and Pragmatic Resources for Turn Management*. In Elinor Ochs, Emanuel Schegloff, and Sandra A. Thompson, eds., Interaction and Grammar, pp.134-184. Cambridge University Press, Cambridge.

61. *Foundation for Intelligent Physical Agents. Specifications*. (1997). Available from http://www.fipa.org

62. Franklin, S., Graesser, A. (1997). *Is it an agent, or just a program?* In Intelligent Agents, III (eds J. P. Miiller, M. Wooldridge and N. R. Jennings), LNAI, Vol.1193, pp.21-36. Springer, Berlin.

63. Gallardo-Paúls, B. (1996). *Análisis conversacional y pragmática del receptor*. Episteme, Valencia.

64. Ganapathibhotla, M., Liu, B. (2008). *Mining Opinions in Comparative Sentences*." In Proceedings of the 22nd International Conference on Computational Linguistics, pp.18-22. Manchester.

65. García, P., Segarra, E., Vidal, E., Galiano, I. (1990). *On the use of the morphic generator grammatical inference methodology in automatic speech recognition*. International Journal of Pattern Recognition and Artificial Intelligence (IJPRIA), Vol.4, pp.667–685.

66. Gee, J.P. (1999). *Introduction to Discourse Analysis*. Routledge.

67. Gelernter, P. (1985). *Generative communication in Linda*. ACM Transactions on Programming Languages and Systems, Vol.7, Issue.1, pp.80-112.

68. Glass J., Flammia, G., Goodine, D., Phillips, M., Polifroni, J., Sakai, S., Seneff, S., Zue, V. (1995). *Multilingual spoken-language understanding in the MIT Voyager system*. Speech Commun., Vol.17, pp.1–18.

69. Göbel, S., Schneider, O., Iurgel, I., Feix, A., Knöpfle, C., and Rettig, A. (2006). *VirtualHuman: Storytelling and Computer Graphics for a Virtual Human Platform*. In Proceedings of the 2nd International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE 2004), pp. 79–88. Springer. Darmstadt, Germany.

70. Goffman, E. (1974). *Frame analysis*. Harper and Row. New York.

71. Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

72. Goodwin, MH. (1990). *He-Said-She-Said: Talk as Social Organization among Black Children*. Indiana University Press, Bloomington.

73. Goodwin. C. (1994). *Professional vision*' American Anthropologist 96, Vol.3, pp.606-33.

74. Gottlieb, K. Windisch, V. (1783). *Briefe über den Schachspieler des Hrn. von Kempelen, nebst drey Kupferstichen die diese berühmte Maschine vorstellen*. Basel, Schweiz.

75. Gravano, A., Hirschber, J. (2009). *Backchannel-Inviting Cues in Task-Oriented Dialogue*. In Proceedings of Interspeech 2009. Brighton.

76. Gravano. A., Hirschberg, J. 2010. *Turn-taking cues in task-oriented dialogue. Computer Speech and Language*. Vol.25, pp.601–634.

77. Grice, H.P. (1982). *Meaning Revisited*, In: N.V. Smith (Ed.), Mutual knowledge, Academic Press, London, pp.223-243.

78. Grice, HP. (1975). *Logic and Conversation*. In Cole P. y J. L. Morgan, pp.41-58.

79. Griol-Barres, D. (2008). *Desarrollo y evaluación de diferentes metodologías para la gestión automática del diálogo*. Ph.D. dissertation (Spanish), Univ. Politécnica de Valencia, Spain, 2008.

80. Grosz, BJ., Sidner CL. (1990). *Plans for Discourse*. In P. R. Cohen, J. Morgan and M. E. Pollack, eds. Intentions in Communication. Cambridge, MA: MIT Press.

81. Grosz, BJ., Sidner, CL. (1986). *Attention, intentions and the structure of discourse*. Computational Linguistics, Vol.12, Issue.3.

82. Gruber, T. (1993). *A translation approach to portable ontology specifications*. In Knowledge Acquisition. Vol. 5, pp.199-199.

83. Gumperz, JJ. (1982). *Discourse Strategies*. Studies in Interactional Sociolinguistics, nº1. Cambridge University Press., Cambridge.

84. Gustafson, J., Boye, J., Fredriksson, M., Johannesson, L., Königsmann, J. (2005). *Providing computer game characters with conversational abilities*. In Proc. Intelligent Virtual Agent (IVA05). Kos, Greece.

85. Hanks, WF. (1990). *Referencial Practice: Language and Lived Space Among the Maya*. Universtiy of Chicago Press. 1990. Chicago.

86. Health, SB. (1983). *Ways with Words: Language, Life and Work in Comunities and Classrooms*. Cambridge: Cambridge University Press.

87. Hershey, J., Movellan, J. (1999). *Using audio-visual synchrony to locate sounds*. In T. K. L. S.A. Solla and K.-R. Mller, editors, Proceedings of 1999 Conference on Advances in Neural Information Processing Systems.

88. Hill, H., Irvine, JT. (1992). *Responsability and Evidence in Oral Discourse*. Cambridge University Press., Cambridge.

89. Hillier FS., Lieberman GJ. (1990). *Introduction to Mathematical Programming*. Mc Graw-Hill. Mexico.

90. Hobbs, JR. Evans, DA. (1980). *Conversation as planned behavior*. Cognitive Science, Vol.4, Issue.4, pp.349-377.

91. International Telecommunication Union (ITU-T) (2003). *One-way Transmission Time*. ITU-T Recommendation G. 114.

92. Jefferson, G. (1989). *Preliminary notes on a possible metric which provides for a standard maximum silence of approximately a second in conversation*. Conversation, pp. 166-196.

93. Jefferson, G. 1984. *Notes on a systematic deployment of the acknowledgement tokens "yeah" and "mmhm"*. Research on Language & Social Interaction. Vol.17, pp.197–216.

94. Jensen, B., Froidevaux, G., Greppin, X,. Lorotte, A., Mayor, L., Meisser, M., Ramel, G., Siegwart, R. (2002). *The interactive autonomous mobile system robox*. In Proceedings of the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems.

95. Joerding, T. (1999). *A temporary user modeling approach for adaptive shopping on the Web*. In: Brusilovsky, P., De Bra, P. (eds.) Proc. of Second Workshop on Adaptive Systems and User Modeling on the World Wide Web, pp.75-79.

96. Johnston, M., Bangalore, S., Vasireddy, G., Stent, A., Ehlen, P., Walker, M., Whittaker, S., Maloor, P. *MATCH: an architecture for multimodal dialogue systems*. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp.376-383. Association for Computational Linguistics, Morristown, NJ, USA.

97. Jokinen, K. (1996). *Cooperative Response Planning in CDM*. Proceedings of the 11th Twente Workshop on Language Technology: Dialogue Management in Natural Language Processing Systems, pp.159-168. Twente, The Netherlands.

98. Jönsson, A. (1997) *A Model for Habitable and Efficient Dialogue Management for Natural Language Interaction*. Natural Language Engineering, Vol.3, pp.103–122.

99. Jurafsky, D., Shriberg, E., Fox, B., Curl, T. 1998. *Lexical, prosodic and syntactic cues for dialog acts*. In: Proceedings of ACL/COLING, Workshop on Discourse Relations and Discourse Markers. pp.114–120.

100. Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., Feiner, S. (2003). *Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality*. In Proceedings of the 5th international conference on Multimodal interfaces, pp.12-19. ACM New York, NY.

101. Kamar, E., Gal, Y., Grosz, BJ. (2009). *Incorporating Helpful Behavior into Collaborative Planning*. AAMAS.

102. Kamar, E., Gal, Y., Grosz, BJ. (2009). *Modeling User Perception of Interaction Opportunities for Effective Teamwork*. SIN09, In Proceedings of IEEE SocialCom.

103. Kempen, G., Hoenkamp, E. (1982). *Incremental sentence generation: Implications for the structure of a syntactic processor*. Proceedings of the Ninth International Conference on Computational Linguistics. Prague.

104. Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

105. Kendon, A. 1967. *Some functions of gaze-direction in social interaction*. Acta Psychologica. Vol.26, pp.22–63.

106. Kilger, A. Finkler, W. (1995). *Incremental generation for real-time applications*. Technical Report RR-95-11, Deutsches Forschungzentrum fÄur KÄunstliche Intelligenz GmbH (DFKI), SaarbrÄucken, Germany.

107. King, M., Maegard, B., Schutz, J., des Tombes, L. (1996). *EAGLES: Evaluation of natural language processing systems*. In Communications of the ACM (39)1, pp.73–79.

108. Koda, T., Maes, P. (1996). *Agents with Faces: The Effect of Personification*. In Proceedings of 5th International Workshop on Robot and Human Communication, pp.189-194. Tsukuba, Japan.

109. Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., Den, Y. 1998. *An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs*. Language and Speech. Vol.41, pp.295–321, Special issue on prosody and conversation.

110. Komatani, K., Rudnicky, A. (2009). *Predicting Barge-in Utterance Errors by using Implicitly Supervised ASR Accuracy and Barge-in Rate per Use*r. Proceedings of the ACL-IJCNLP, Suntec, Singapore, pp.89-92.

111. Kulick, D. (1992). *Language Shift and Cultural Reproduction: Socialization, Self, and Syncretism in a Papua New Guinean Village*. Cambridge University Press, Cambridge.

112. Lamel, L., Bennacef, SK., Rosset, S., Devillers, L., Foukia, S., Gangolf, JJ., Gauvain, JL. (1997). *The LIMSI RailTel system: Field trial of a telephone service for rail travel information*. Speech Commun., Vol. 23, pp. 67–82.

113. Lemon, O., Bracy, A., Gruenstein, A., Peters, S. (2001). *The WITAS multi-modal dialogue system I*. In Proc. European Conf. on Speech Communication and Technology, pp.1559–1562. Aalborg, Denmark.

114. Lenzmann, B., Wachsmuth, I. (1997). *Contract-Net-Based Learning in a User-Adaptive Interface Agency*. In Weiss, G. (ed.): Distributed Articial Intelligence Meets Machine Learning: Learning in Multi-Agent Environments, pp. 202-222. Berlin: Springer.

115. Levelt, WJM. (1989). *Speaking*. MIT Press., Cambridge MA. 1989

116. Levinson, SC. (1983). *Pragmatics*. Cambridge: Cambridge Univ. Press.

117. Levinson, SC. (1992). *Primer for the field investigation of spatial description and conception*. Pragmatics, Vol.2, Issue.1, pp.5–47.

118. Linell, P. (1998). *Approaching dialogue: talk, interaction and contexts in dialogical perspectives*. Benjamins, Amsterdam.

119. Liu, X., Zhao, Y., Pi, X., Liang, L., Nefian, AV. (2002). *Audio-visual continuous speech recognition using a coupled hidden Markov model*. In Proc. Int. Conf. Spoken Lang. Processing, pp.213-216. Denver, CO.

120. Löckelt, M., Pfleger, N. (2005). *Multi-Party Interaction With Self-Contained Virtual Characters*. Proceedings of the Ninth Workshop on the Semantics and Pragmatics of Dialogue (DIALOR'05) (Poster), LORIA, Nancy, France.

121. Löckelt. M. (2008). *A Flexible and Reusable Framework for Dialogue and Action Management in Multi-Party Discourse*. PhD-Thesis, Universität des Saarlandes.

122. López-Cózar, R., Araki, M. (2005). *Spoken, Multilingual and Multimodal Dialogue Systems: Development and Assessment*. John Wiley & Sons Publishers.

123. Mariani, J. (2002). *Technolangue: language technology. Language Technologies: Technolangue Action*. Presentation. In: LREC 2002 International Strategy Panel 17. Las Palmas, Spain.

124. Mauldin, ML. (1994). *CHATTERBOTS, TINYMUDS and the Turing Test: Entering the Loebner Prize Competition*. In AAAI-94 Proceedings.

125. McTear, M. (1999). *Using the CSLU Toolkit for Practicals in Spoken Dialogue Technology*. In Proceedings of ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education, London, UK.

126. McTear, M. (2002). *Spoken dialogue technology: Enabling the conversational user interface*. ACM Comput. Surv., Vol.34, pp.90–169.

127. McTear, M. (2004). *Spoken dialogue technology: toward the conversational user interface*. Springer-Verlag, London.

128. Meza, I., Perez, E., Salinas, L., Aviles, H., Pineda, LA. (2010). *A Multimodal Dialogue System for Playing the Game Guess the card*. Procesamiento del Lenguaje Natural, Revista, nº44, pp.131-138.

129. Moubaiddin, A., Obeid, N. (2008). *Partial information basis for agent-based collaborative dialogue*. Applied Intelligence, Vol.30, Issue.2, pp.142-167.

130. Mushin, I., Stirling, L., Fletcher, J., Wales, R. 2003. *Discourse structure, grounding, and prosody in task-oriented dialogue*. Discourse Processes. Vol.35, pp.1–31.

131. Norman, DA. (1988). *The psychology of everyday things*. Basic books, New York.

132. Novick, D., Sutton, S. 1994. *An empirical model of acknowledgment for spoken-language systems*. In: Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics, Morristown, NJ, USA. pp.96–101.

133. Oberle, D., Ankolekar, A., Hitzler, P., Cimiano, P., Sintek, M., Kiesel, M., Mougouie, B., Vembu, S., Baumann, S., Romanelli, M., Buitelaar, P., Engel, R., Sonntag, D., Reithinger, N., Loos, B., Porzel, R., Zorn, HP., Micelli, V., Schmidt, C., Weiten, M., Burkhardt, F., Zhou, J. (2006). *DOLCE ergo SUMO: On Foundational and Domain Models in SWIntO (SmartWeb Integrated Ontology)*. Technical report, AIFB, University of Karlsruhe.

134. Ochs, E. (1992). *Indexing gender*. In A. Duranti and C. Goodwin (Eds.) Rethinking Context: Language as an Interactive Phenomenon, pp.223-58. Cambridge University Press. Cambridge.

135. Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., Ferro, D. (2000). *Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions*. Human-Computer Interaction, Vol.15, Issue.4, pp.263-322.

136. Oviatt, SL. (1996). *Multimodal Interfaces for Dynamic Interactive Maps*. CHI 1996, pp.95-102.

137. Pallet, DS., Fiscus, JG., Fisher, WM, Garofolo, JS., Lund, BS., Martin, A., Przybocki, M.A. (1995). *The 1994 Benchmark Tests for the ARPA Spoken Language Program*. In Proceedings of ARPA Workshop on Spoken Language Technology.

138. Paterno, F. (2000). *Model-based design of interactive applications*. Intelligence, Vol.11, Issue.4, pp.26-38. ACM New York, NY, USA.

139. Pfleger, N. (2007). *Context-based Multimodal Interpretation: An Integrated Approach to Multimodal Fusion and Discourse Processing*. Ph.D. dissertation, Universtität des Saarlandes, Germany.

140. Pfleger, N. (2007). *Context-based Multimodal Interpretation: An Integrated Approach to Multimodal Fusion and Discourse Processing*. Ph.D. dissertation, Universtität des Saarlandes, Germany.

141. Poppe, R., Truong, KP., Reidsma, D., Heylen, D. 2010. *Backchannel Strategies for Artificial Listeners*. In: J. Allbeck et al. (Eds.). IVA 2010, Springer-Verlag berling Heielberg. pp.146-158.

142. Poyatos, F. (1980). *Interactive functions and limitations of verbal and non verbal behavior in natural conversations*. Semiotica, Vol.30, Issue.3/4, pp.211-244.

143. Poyatos, F. (1994). *Paralenguaje, kinésica e interacción*. La comunicación no verbal, Vol.2. Itsmo, Madrid.

144. Raux, A., Eskenazi, M. 2008. *Optimizing end pointing thresholds using dialogue features in a spoken dialogue system*. In: SIGdial, Columbus,OH.

145. Reithinger, N., Kipp, M. (1998). *Large scale dialogue annotation in Verbmobil*. In Workshop Proceedings of ESSLLI 98.

146. Rich, C., Sidner, CL., and Lesh, N. (2001). *COLLAGEN: Applying collaborative discourse theory to human-computer interaction*. AI Magazine, Vol.22, nº4, pp.15–25. Special Issue on Intelligent User Interfaces.

147. Roberts, GL., Bavelas, JB. (1996). *The Communicative Dictionary: A Collaborative Theory of Meaning*. In Stewart, J. (Eds) Beyond the Symbol Model, pp.135--160. State University of New York Press, Albany.

148. Rogozan, A., Deléglise, P. (1998). *Adaptive Fusion of Acoustic and Visual Sources for Automatic Speech Recognition*. In Speech Communication Journal, Vol.26 Issue.1-2, pp.149-161.

149. Sacks, H., Schegloff EA., Jefferson. G. (1974). *A simplest systematics for the organization of turn-taking in conversation*. Language, Vol.50, pp.696-735.

150. Sadek, D. (1999). *Design considerations on dialogue systems: from theory to technology*. In Procs. of the ESCA workshop on Interactive Dialogue Systems, Kloster Irsee, Germany.

151. Saussure, F. (1916). Curso de Lingüística General, Alianza, Edición de Tullio de Mauro, Madrid.

152. Schaffer, D. 1983. *The role of intonation as a cue to turn taking in conversation*. Journal of Phonetics. vol.11, pp.243–257.

153. Schegloff, E. 1982. *Discourse as an interactional achievement: some uses of "uhh, uh" and other things that come between sentences*. In: Analyzing Discourse: Text and Talk. pp.71–93.

154. Schegloff, EA. (2007). *Sequence Organization in Interaction*. Cambridge University Press.

155. Schieffelin, BB. (1990). *The Give and Take of Everyday: Language Socialization of Kaluli Children*. Cambridge University Press., Cambridge.

156. Schiffrin, D. (1994). *Approaches to Discourse*. Blackwell, Cambridge, Mass.

157. Schlangen, D. 2006. *From reaction to prediction: Experiments with computational models of turn-taking*. In: Proceedings of Interspeech.

158. Schwartz, D., Sterling, L., Mayland, E. (1991). *FLiPSide Blackboard: A Financial Logic Programming System for Distributed Expertise.* Proc. First Int'l Conf. AI Application on Wall Street.

159. Searle, J. (1969). *Speech Acts*. Cambridge University Press.

160. Segarra, E., Hurtado, L. (1997). *Construction of Language Models using Morfic Generator Grammatical Inference MGGI Methodology*. In Proc. of European Conference on Speech Communications and Technology (Eurospeech'97), págs. 2695–2698, Rodas (Grecia).

161. Seneff, S., Wang, C., Chao, CY. (2007). *Spoken Dialogue Systems for Language Learning*. NAACL-HLT 2007.

162. Sinclair, J., Coulthard, M. (1975). *Towards an Analysis of Discourse*. Oxford: Oxford University Press.

163. Stalnaker, R. (1978). *Assertion*. Syntax and Semantics, Vol.9, pp.315–322.

164. Swartout, W., Paris, C., Moore, J. (1991). *Explanations in Knowledge Systems: Design for Explainable Expert Systems*. IEEE Expert: Intelligent Systems and Their Applications, Vol.6, Issue.3, pp.58-64. IEEE Educational Activities Department Piscataway, NJ, USA.

165. Swerts, M. (1998). *Filled pauses as markers of discourse structure. Journal of Pragmatics*. Vol.30, Issue.4, pp.485-496.

166. Tannen, D. (1989). *Interpreting interruption in conversation*. Papers from the 25th annual regional meeting of the Chicago Linguistic Society, Part II: Parasession on Language in Context, pp.266-87. Chicago Linguistic Society, Chicago.

167. Tannen, D. (2001). *You Just Don't Understand: Women and Men in Conversation*. New paperback edition: New York: Quill.

168. Thomson, B., Young, S. (2010). *Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems*. Computer Speech and Language, Vol.24, Issue.4, pp.562-588.

169. Thórisson, K. R. (2002). *Natural Turn-Taking Needs No Manual: Computational Theory and Model, from Perception to Action*. Multimodality in Language and Speech Systems, pp.173–207. Kluwer Academic Publishers, Dordrecht, The Netherlands.

170. Traum, D. Larsson, S. (2003). *The Information State Approach to Dialogue Management*. In Smith and Kuppevelt (eds.): Current and New Directions in Discourse & Dialogue, pp. 325-353. Kluwer Academic Publishers.

171. Traum, D., Bos, J., Cooper, R., Larson S., Lewin, I., Mathesson, C., Poesio, M. (1999). *A model of Dialogue Moves and Information State Revision*. Trindi Technical Report D2.1.

172. Turunen, M., Hakulinen, J. (2000). *Mailman - a Multilingual Speech-only E-mail Client based on an Adaptive Speech Application Framework*. Proceedings of Workshop on Multi-Lingual Speech Communication (MSC 2000), pp.7-12.

173. Valle-Agudo, D., Calle-Gómez, FJ., Cuadra-Fernández, D., Rivero-Espinosa, J. (2009). *Breaking of the Interaction Cycle: Independent Interpretation and Generation for Advanced Dialogue Management*. Human-Computer Interaction. Novel Interaction Methods and Techniques. Lecture Notes in Computer Science, Vol.5611/2009, pp.674-683. Springer Berlin / Heidelberg.

174. Valle-Agudo, D., Calle-Gómez, FJ., Martínez-Fernández, P. (2006). *Enfoque Metodológico para Incorporar Conocimiento de Dominio a Sistemas de Diálogo Intencionales*. Revista Española para el procesamiento del lenguaje natural, nº37, pp.25-32. Zaragoza, Spain.

175. Valle-Agudo, D., Rivero-Espinosa, J., Calle-Gómez, FJ., Cuadra-Fernández, D. (2007). *Conocimiento Circunstancial en Sistemas de Interacción Natural*. Simposio de Computación Ubicua e Inteligencia Ambiental, pp.177-184. Thomson, Zaragoza, Spain, September.

176. Vanderheiden, GC., Zimmermann, G. (2005). *Use of user interface sockets to create naturally evolving intelligent environments*. 11th International Conference on Human-Computer Interaction, Caesars Palace, Las Vegas, Nevada USA.

177. Voice Extensible Markup Language (VoiceXML) 2.1. W3C Recommendation. W3C. (2007). http://www.w3.org/TR/2007/REC-voicexml21-20070619/

178. Wahlster, W. (2006). *SmartKom: foundations of multimodal dialogue systems*. Springer.

179. Wahlster, W., Marburger, H., Jameson, A., Busemann, S. 1983. *Over-answering yes-no questions: Extended responses in a nl interface to a vision system*. In International Joint Conference Artificial Intelligence, pp.643-646.

180. Walker, M., Litman, D., Kamm, C., Abella, A. (1997). *PARADISE: A general framework for evaluating spoken dialogue agents*. In Proceedings of the 35th Annual Meeting of the Association of Computational Linguistics. pp. 271–280.

181. Walton, KL. (1983). *Fiction, fiction-making, and styles of fictionality*. Philosophy and Literature, Vol.8, pp.78-88.

182. Walton, KL. (1990). *Mimesis as make-believe: On the foundations of the representational arts*. Hardvard University Press, Cambridge MA.

183. Wang, K. (2000). *Implementation of a multimodal dialog system using extensible markup language*. Proc. ICSLP-2000. Beijing, China.

184. Wang, K. (2002). *SALT: an XML application for web-based multimodal dialog management*. In NLPXML '02: Proceedings of the 2nd workshop on NLP and XML, pp.1–8. Association for Computational Linguistics, Morristown, NJ, USA.

185. Ward W., Pellom, B. (1999). *The CU Communicator system*. IEEE Workshop Automatic Speech Recognition and Understanding, pp. 1999.

186. Ward, N., Tsukahara, W. 2000. *Prosodic features which cue back-channel responses in English and Japanese*. Journal of Pragmatics. vol.32, pp.1177–1207.

187. Wasinger, R., Stahl, C., Krueger, A. (2003). *Robust speech interaction in a mobile environment through the use of multiple and different media input types*. In EUROSPEECH-2003, pp.1049-1052.

188. Weiser, M. (1991). *The Computer for the Twenty-First Century*. Scientific American, Vol.265, nº3, pp.94-104.

189. Weitzenbaum, J. (1976). *Computer Power and Human Reason*. San Francisco: W.H. Freeman.

190. Wennerstrom, A., Siegel, AF. 2003. *Keeping the floor in multi-party conversations: intonation, syntax, and pause*. Discourse Processes. Vol. 36, pp.77–107.

191. Wierzbicka, A. (1987). *English Speech Act Verbs: A semantic dictionary*. Academic Press in Sydney, Orlando.

192. Williams, JD. (2008). *Demonstration of a POMDP voice dialer*. In Proc Demonstration Session ACL-HLT.

193. Winograd, T. (1971). *Procedures as a Representation for Data in a Computer Program for Understanding Natural Language*. National Technical Information Service, Springfield.

194. Wooldridge, M. (2002). *An Introduction to Multi-Agent Systems*. John Wiley and Sons Limited: Chichester.

195. Xu, W., Rudnicky, A. (2000). *Language modeling for dialog system*. In Proceedings of ICSLP 2000, Beijing, China.

196. Yngve, V. 1970. *On getting a word in edgewise*. In: Proceedings of the Sixth Regional Meeting of the Chicago Linguistic Society. Vol.6, pp.657–677.

197. Young, SL., Hauptmann, AG., Ward, W H., Smith, ET., Werner, P. (1989). *High level knowledge sources in usable speech recognition systems*. Communications of the ACM, Vol.32, Issue.2, pp.183-194, ACM New York, NY, USA.

198. Zimmerman, D., West, C. (1975). *Sex Roles, Interruptions and Silences in Conversations*. In Thorne, Barrie/Henley, Nancy (eds.), Language and Sex: Difference and Dominance. Newbury House, Rowley, MA.

199. Zue, V., Seneff, S., Glass, JR., Polifroni, J., Pao, C., Hazen, TJ., Hetherington, L. (2000). *JUPITER: A Telephone-Based Conversational Interface for Weather Information*. IEEE Transactions on Speech and Audio Processing, Vol.8, nº1.